

迁移中心 MgC

用户指南

文档版本 22
发布日期 2025-02-14



版权所有 © 华为技术有限公司 2025。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

安全声明

漏洞处理流程

华为公司对产品漏洞管理的规定以“漏洞处理流程”为准，该流程的详细内容请参见如下网址：

<https://www.huawei.com/cn/psirt/vul-response-process>

如企业客户须获取漏洞信息，请参见如下网址：

<https://securitybulletin.huawei.com/enterprise/cn/security-advisory>

目录

1 新版	1
1.1 权限管理	1
1.1.1 创建用户并授权使用 MgC	1
1.1.2 MgC 自定义策略	5
1.1.3 委托权限说明	6
1.1.4 IAM 用户权限	15
1.2 配置管理	16
1.2.1 项目管理	16
1.2.2 凭证管理	18
1.3 调研评估	20
1.3.1 IDC 上云成本对比	20
1.3.2 跨云迁移成本分析（基于凭证）	27
1.3.3 跨云迁移成本分析（基于账单）	30
1.3.4 问卷调查	35
1.3.5 存储迁移成本调研	36
1.4 资源采集	36
1.4.1 在线调研	37
1.4.2 内网采集	45
1.4.3 导入 RVTools 资源	49
1.4.4 导入阿里云资源清单	51
1.4.5 导入应用关联文件	55
1.4.6 导入工具采集结果	55
1.4.7 手动添加资源	56
1.4.8 采集主机性能	59
1.4.9 关联应用	61
1.5 应用管理	62
1.5.1 创建应用	62
1.5.2 关联资源	63
1.5.3 手动添加资源	63
1.5.4 采集主机性能	65
1.6 迁移准备	67
1.6.1 部署 MgC Agent（原 Edge）	68
1.6.2 采集元数据	74

1.6.2.1 创建元数据连接.....	74
1.6.2.2 创建并运行元数据同步任务.....	83
1.6.2.3 创建并运行数据湖元数据同步任务.....	85
1.6.2.4 查看元数据.....	87
1.6.3 血缘采集.....	88
1.6.3.1 血缘采集概述.....	88
1.6.3.2 导出作业.....	89
1.6.3.3 创建血缘采集任务.....	91
1.6.3.4 查看血缘关系.....	93
1.7 迁移方案设计.....	93
1.7.1 绑定目的端.....	94
1.7.2 规格评估.....	95
1.7.3 购买资源.....	99
1.7.4 迁移风险评估.....	100
1.8 迁移方案配置.....	102
1.8.1 配置主机迁移方案.....	102
1.8.2 配置存储批量迁移方案.....	110
1.8.3 配置文件存储批量迁移方案.....	112
1.9 迁移集群.....	116
1.9.1 创建集群.....	116
1.9.2 集群管理.....	120
1.9.3 计费说明.....	122
1.9.4 集群状态说明.....	122
1.9.5 资源类型和安全组配置说明.....	123
1.10 迁移 workflow.....	127
1.10.1 workflow 配额限制说明.....	127
1.10.2 创建主机迁移 workflow.....	127
1.10.3 创建可用区迁移 workflow.....	136
1.10.4 创建存储迁移 workflow.....	137
1.10.5 创建对象存储批量迁移 workflow.....	150
1.10.6 创建文件存储批量迁移 workflow.....	155
1.10.7 添加阶段/步骤.....	158
1.11 大数据数据迁移（MaxCompute 迁移到 DLI）.....	159
1.11.1 大数据迁移概述.....	159
1.11.2 使用须知.....	160
1.11.3 准备工作.....	161
1.11.4 创建源端连接.....	171
1.11.5 创建目的端连接.....	172
1.11.6 创建迁移任务.....	174
1.11.6.1 元数据全量迁移.....	174
1.11.6.2 元数据增量感知.....	178
1.11.6.3 全量数据迁移.....	182

1.11.6.4 增量数据迁移.....	188
1.11.7 查看并导出执行结果.....	194
1.11.8 自定义参数说明.....	196
1.12 大数据校验.....	198
1.12.1 大数据校验概述.....	198
1.12.2 准备工作.....	201
1.12.3 创建源端连接.....	202
1.12.3.1 创建源端云服务连接.....	202
1.12.3.2 创建源端组件连接.....	204
1.12.4 创建表组并添加数据表.....	211
1.12.5 创建目的端连接.....	213
1.12.5.1 创建执行机连接.....	214
1.12.5.2 创建目的端云服务连接.....	215
1.12.5.3 创建目的端组件连接.....	217
1.12.6 创建统计任务并执行校验.....	222
1.12.7 查看并导出校验结果.....	247
1.12.8 自定义参数说明.....	249
2 旧版.....	251
2.1 权限管理.....	251
2.1.1 创建用户并授权使用 MgC.....	251
2.1.2 MgC 自定义策略.....	255
2.1.3 委托权限说明.....	256
2.1.4 IAM 用户权限.....	265
2.2 配置管理.....	266
2.2.1 项目管理.....	267
2.2.2 凭证管理.....	268
2.3 上云调研.....	270
2.3.1 上云成本分析（凭证）.....	270
2.3.2 上云成本分析（账单）.....	273
2.3.3 问卷调研.....	278
2.3.4 存储迁移成本调研.....	279
2.4 资源采集.....	279
2.4.1 复杂项目.....	279
2.4.1.1 公网发现与采集.....	280
2.4.1.2 内网发现与采集.....	290
2.4.1.3 导入应用关联文件.....	292
2.4.1.4 导入工具采集结果.....	293
2.4.1.5 导入阿里云资源清单.....	294
2.4.1.6 导入 RVTools 资源.....	298
2.4.2 简单项目.....	301
2.4.2.1 公网发现与采集.....	301
2.4.2.2 内网发现与采集.....	310

2.4.2.3 导入阿里云资源清单.....	313
2.4.2.4 导入 RVTools 资源.....	317
2.4.3 手动添加资源.....	320
2.4.4 采集主机性能.....	323
2.4.5 关联应用.....	324
2.5 应用管理.....	325
2.5.1 创建应用.....	325
2.5.2 关联资源.....	326
2.5.3 手动添加资源.....	327
2.5.4 采集主机性能.....	330
2.5.5 查看应用关联分析结果.....	331
2.6 大数据血缘.....	333
2.6.1 部署迁移工具 Edge.....	333
2.6.2 采集元数据.....	333
2.6.2.1 创建元数据连接.....	333
2.6.2.2 创建并运行元数据同步任务.....	340
2.6.2.3 创建并运行数据湖元数据同步任务.....	342
2.6.2.4 查看元数据.....	344
2.6.3 血缘采集.....	346
2.6.3.1 血缘采集概述.....	346
2.6.3.2 导出作业.....	346
2.6.3.3 创建血缘采集任务.....	348
2.6.3.4 查看血缘关系.....	350
2.7 迁移方案设计.....	351
2.7.1 绑定目的端.....	351
2.7.2 规格评估.....	352
2.7.3 购买资源.....	356
2.7.4 迁移风险评估.....	357
2.8 迁移方案配置.....	359
2.8.1 配置主机迁移方案.....	359
2.8.2 配置存储批量迁移方案.....	367
2.8.3 配置文件存储批量迁移方案.....	369
2.9 迁移集群.....	373
2.9.1 创建集群.....	373
2.9.2 集群管理.....	377
2.9.3 计费说明.....	379
2.9.4 集群状态说明.....	379
2.9.5 资源类型和安全组配置说明.....	380
2.10 迁移 workflow.....	384
2.10.1 workflow 配额限制说明.....	384
2.10.2 创建主机迁移 workflow.....	384
2.10.3 创建可用区迁移 workflow.....	390

2.10.4 创建存储迁移工作流.....	392
2.10.5 创建对象存储批量迁移工作流.....	405
2.10.6 创建文件存储批量迁移工作流.....	408
2.10.7 添加阶段/步骤.....	411
2.11 大数据数据迁移.....	413
2.11.1 使用须知.....	413
2.11.2 准备工作.....	413
2.11.3 创建源端连接.....	423
2.11.4 创建目的端连接.....	425
2.11.5 迁移实施.....	427
2.11.5.1 元数据全量迁移.....	427
2.11.5.2 元数据增量感知.....	431
2.11.5.3 全量数据迁移.....	435
2.11.5.4 增量数据迁移.....	441
2.11.6 查看并导出执行结果.....	446
2.11.7 自定义参数说明.....	448
2.12 大数据校验.....	450
2.12.1 大数据校验概述.....	450
2.12.2 准备工作.....	452
2.12.3 创建源端连接.....	453
2.12.3.1 创建源端云服务连接.....	453
2.12.3.2 创建源端组件连接.....	454
2.12.4 创建表组并添加数据表.....	459
2.12.5 创建目的端连接.....	462
2.12.5.1 创建执行机连接.....	462
2.12.5.2 创建目的端云服务连接.....	463
2.12.5.3 创建目的端组件连接.....	465
2.12.6 创建统计任务并执行校验.....	468
2.12.7 查看并导出校验结果.....	491
2.12.8 自定义参数说明.....	492

1 新版

1.1 权限管理

1.1.1 创建用户并授权使用 MgC

如果您需要对您所拥有的MgC进行精细的权限管理，您可以使用[统一身份认证服务](#)（Identity and Access Management，简称IAM），通过IAM，您可以：

- 根据企业的业务组织，在您的华为云账号中，给企业中不同职能部门的员工创建IAM用户，让员工拥有唯一安全凭证，并使用MgC资源。
- 根据企业用户的职能，设置不同的访问权限，以达到用户之间的权限隔离。
- 将MgC资源委托给更专业、高效的其他华为云账号或者云服务，这些账号或者云服务可以根据权限进行代运维。

如果华为云账号已经能满足您的要求，不需要创建独立的IAM用户，您可以跳过本章节，不影响您使用MgC服务的其它功能。

本章节为您介绍对用户授权的方法，操作流程如[示例流程](#)所示。

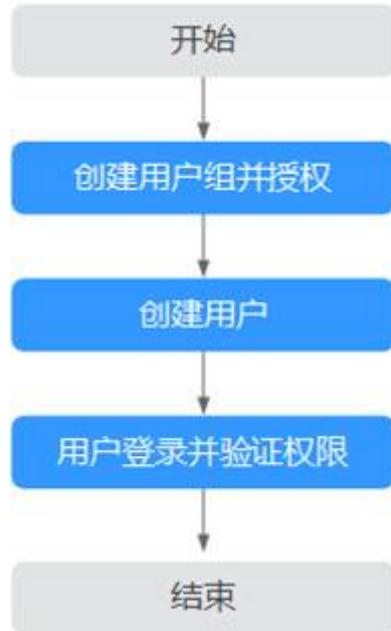
前提条件

给用户组授权之前，请您了解用户组可以添加的MgC系统策略，并结合实际需求进行选择。

MgC支持的系统策略及策略间的对比，请参见：[MgC系统策略](#)。若您需要对除MgC之外的其它服务授权，IAM支持服务的所有策略请参见[权限策略](#)。

示例流程

图 1-1 给用户授权 MgC 权限流程



1. 创建用户组并授权

- **系统策略**: 在IAM控制台创建用户组，并根据**MgC系统策略说明**以及实际的权限要求，为用户组授权MgC系统策略，授权范围方案选择“所有资源”。

表 1-1 MgC 系统策略说明

策略名称	描述	策略类别	策略内容
MgC FullAccess	迁移中心管理员权限，拥有操作MgC的所有权限。	系统策略	MgC FullAccess策略内容
MgC ReadOnlyAccess	迁移中心只读权限，仅能查看MgC资源，无法进行操作。	系统策略	MgC ReadOnlyAccess策略内容
MgC DiscoveryAccess	迁移中心资源发现操作权限，拥有操作资源发现功能的权限和只读权限。	系统策略	MgC DiscoveryAccess策略内容
MgC AssessAccess	迁移中心评估操作权限，拥有操作评估功能、资源发现功能的权限和只读权限。	系统策略	MgC AssessAccess策略内容
MgC MigrateAccess	迁移中心迁移操作权限，拥有操作迁移功能、评估功能、资源发现功能的权限和只读权限。	系统策略	MgC MigrateAccess策略内容

策略名称	描述	策略类别	策略内容
MgC AppDiscovery Access	迁移中心资源采集操作权限，拥有操作资源采集功能、资源发现功能的权限和只读权限。	系统策略	MgC AppDiscovery Access策略内容
MgC MrrAccess	迁移中心业务验证操作权限，拥有业务验证功能的权限和只读权限。	系统策略	MgC MrrAccess策略内容

- **自定义策略：**如果IAM用户只需要拥有迁移中心 MgC部分操作权限，则使用自定义策略，参见[MgC自定义策略](#)。
2. **创建用户并加入用户组**
在IAM控制台创建用户，并将其加入[1.创建用户组并授权](#)中创建的用户组。

须知

当IAM用户首次使用MgC的特定功能时，需要通过以下两种方式授予委托权限：

- **推荐方式：**请管理员登录MgC控制台，进入相关功能界面，同意相关委托权限。
- **备选方式：**请管理员登录“统一身份认证服务 IAM”服务，为当前IAM用户配置创建委托所需的权限，确保其拥有MgC相关系统策略以及"iam:agencies:createAgency"、"iam:permissions:grantRoleToAgency"、"iam:roles:createRole"和"iam:roles:updateRole"的权限。

3. **用户登录**并验证权限

新创建的用户登录控制台，切换至授权区域，验证权限：

- 在“**服务列表**”中选择迁移中心 MgC，进入MgC主界面，根据您授予的权限可以进行对应操作，表示授予的权限已生效。
- 在“**服务列表**”中选择除迁移中心 MgC外的任一服务，若提示权限不足，表示授予的权限已生效。

MgC FullAccess 策略内容

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Action": [
        "mgc:*",
        "iam:agencies:listAgencies",
        "iam:roles:listRoles",
        "iam:quotas:listQuotas",
        "iam:permissions:listRolesForAgency"
      ],
      "Effect": "Allow"
    }
  ]
}
```

MgC ReadOnlyAccess 策略内容

```
{
  "Version": "1.1",
```

```
"Statement": [  
  {  
    "Effect": "Allow",  
    "Action": [  
      "mgc:*:query*"  
    ]  
  }  
]
```

MgC DiscoveryAccess 策略内容

```
{  
  "Version": "1.1",  
  "Statement": [  
    {  
      "Effect": "Allow",  
      "Action": [  
        "mgc:*:query*",  
        "mgc:*:discovery"  
      ]  
    }  
  ]  
}
```

MgC AssessAccess 策略内容

```
{  
  "Version": "1.1",  
  "Statement": [  
    {  
      "Action": [  
        "mgc:*:query*",  
        "mgc:*:discovery",  
        "mgc:*:assess",  
        "iam:agencies:listAgencies",  
        "iam:roles:listRoles",  
        "iam:quotas:listQuotas",  
        "iam:permissions:listRolesForAgency"  
      ],  
      "Effect": "Allow"  
    }  
  ]  
}
```

MgC MigrateAccess 策略内容

```
{  
  "Version": "1.1",  
  "Statement": [  
    {  
      "Action": [  
        "mgc:*:query*",  
        "mgc:*:discovery",  
        "mgc:*:assess",  
        "mgc:*:migrate",  
        "iam:agencies:listAgencies",  
        "iam:roles:listRoles",  
        "iam:quotas:listQuotas",  
        "iam:permissions:listRolesForAgency"  
      ],  
      "Effect": "Allow"  
    }  
  ]  
}
```

MgC AppDiscoveryAccess 策略内容

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "mgc:*:query*",
        "mgc:*:discovery",
        "mgc:*:appdiscovery"
      ]
    }
  ]
}
```

MgC MrrAccess 策略内容

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Action": [
        "mgc:*:query*",
        "mgc:mrr:query",
        "mgc:mrr:update",
        "mgc:mrr:export",
        "mgc:mrr:import",
        "mgc:mrr:upgrade",
        "mgc:mrr:delete",
        "mgc:mrr:check"
      ],
      "Effect": "Allow"
    }
  ]
}
```

1.1.2 MgC 自定义策略

如果系统预置的MgC权限，不满足您的授权要求，可以创建自定义策略。

目前华为云支持以下两种方式创建自定义策略：

- 可视化视图创建自定义策略：无需了解策略语法，按可视化视图导航栏选择云服务、操作、资源、条件等策略内容，可自动生成策略。
- JSON视图创建自定义策略：可以在选择策略模板后，根据具体需求编辑策略内容；也可以直接在编辑框内编写JSON格式的策略内容。

具体创建步骤请参见：[创建自定义策略](#)。本章为您介绍常用的MgC自定义策略样例。

MgC 自定义策略样例

- 平台采集、资源管理

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "mgc:*:query*",
        "mgc:*:discovery"
      ]
    }
  ]
}
```

- ```
 }
]
}
```
- TCO评估、主机评估

```
{
 "Version": "1.1",
 "Statement": [
 {
 "Action": [
 "mgc*:query*",
 "mgc*:discovery",
 "mgc*:assess",
 "iam:agencies:listAgencies",
],
 "Effect": "Allow"
 }
]
}
```
  - 应用关联分析

```
{
 "Version": "1.1",
 "Statement": [
 {
 "Effect": "Allow",
 "Action": [
 "mgc*:query*",
 "mgc*:discovery",
 "mgc*:appdiscovery"
]
 }
]
}
```
  - 迁移 workflow

```
{
 "Version": "1.1",
 "Statement": [
 {
 "Effect": "Allow",
 "Action": [
 "mgc*:query*",
 "mgc*:discovery",
 "mgc*:assess",
 "mgc*:migrate",
 "iam:agencies:listAgencies"
]
 }
]
}
```

### 1.1.3 委托权限说明

#### 概述

当您启用特定功能时，需要您确认并同意授予必要的权限给MgC，以便我们能够为您提供完整的服务。本节为您介绍需要进行权限委托的场景以及自定义策略详情。

委托授权策略说明：

- **创建自定义策略**：如果您尚未设定任何自定义的权限委托策略，系统会自动为您创建一个标准的自定义策略。
- **更新自定义策略**：对于已经存在的自定义策略，系统将进行完整性检查。若发现策略内容不全面或缺少必要元素，系统将自动进行补充。

## 跨可用区迁移

| 场景     | 委托对象 | 自定义策略名称                        | 细粒度最小使用权限                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|--------|------|--------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 跨可用区迁移 | MgC  | MgC<br>AzMigrationAgencyPolicy | ecs:cloudServers:showServer ( 查询云服务器详情 )<br>ecs:flavors:get ( 查询云服务器规格 )<br>ecs:cloudServerFlavors:get ( 查询云服务器规格详情和扩展信息列表 )<br>ecs:cloudServerQuotas:get ( 查询租户配额 )<br>ecs:servers:list ( 查询云服务器列表 )<br>ecs:cloudServers:list ( 查询云服务器详情列表 )<br>ecs:servers:stop ( 关闭云服务器 )<br>ecs:cloudServers:listServerInterfaces ( 查询云服务器网卡信息 )<br>ecs:cloudServers:createServers ( 创建云服务器 )<br>ecs:cloudServers:listServerBlockDevices ( 查询弹性云服务器磁盘信息 )<br>ecs:cloudServerNics:update ( 云服务器网卡配置私有IP )<br>ecs:availabilityZones:list ( 查询可用区列表 )<br>ecs:servers:start ( 启动云服务器 )<br>ecs:cloudServers:changeNetworkInterface ( 更新云服务器指定网卡属性 )<br>ecs:serverInterfaces:get ( 查询云服务器网卡 )<br>ecs:cloudServers:get ( 查询云服务器详情 )<br>vpc:publicIps:create ( 创建弹性公网IP )<br>vpc:publicIps:update ( 更新弹性公网IP )<br>vpc:subnets:get ( 查询子网列表或详情 )<br>vpc:networks:get ( 查询网络列表或详情 )<br>vpc:publicIps:list ( 查询弹性公网IP )<br>vpc:publicIps:get ( 查询弹性公网IP详情 )<br>vpc:ports:get ( 查询端口列表或详情 )<br>vpc:ports:delete ( 删除端口 ) |

| 场景 | 委托对象 | 自定义策略名称 | 细粒度最小使用权限                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|----|------|---------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    |      |         | vpc:ports:update (更新端口)<br>vpc:ports:create (创建端口)<br>evs:types:get (查询云硬盘类型)<br>evs:volumes:list (查询云硬盘列表)<br>cbr:vaults:get (查询指定存储)<br>cbr:vaults:list (查询存储库列表)<br>cbr:vaults:create (创建存储库)<br>cbr:vaults:addResources (添加资源)<br>cbr:vaults:backup (执行备份)<br>cbr:backups:list (查询备份列表)<br>cbr:tasks:list (查询任务列表)<br>cbr:tasks:get (查询单个任务)<br>cbr:backups:delete (删除备份)<br>cbr:backups:get (查询指定备份)<br>cbr:vaults:delete (删除存储库)<br>ims:wholeImages:create (制作整机镜像)<br>ims:images:list (查询镜像列表)<br>ims:images:delete (删除镜像)<br>ims:images:get (查询镜像详情)<br>ims:serverImages:create (制作镜像) |

## 上云成本分析

| 场景         | 委托对象 | 自定义策略名称                | 细粒度最小使用权限                                                                                                                  |
|------------|------|------------------------|----------------------------------------------------------------------------------------------------------------------------|
| 创建上云成本分析任务 | MgC  | MgC<br>TcoAgencyPolicy | ecs:cloudServerFlavors:get (查询云服务器规格详情和扩展信息列表)<br>evs:types:get (查询云硬盘类型)<br>ims:*.get* (ims查询权限)<br>ims:*.list* (ims查询权限) |

## 规格评估

| 场景   | 委托对象 | 自定义策略名称                             | 细粒度最小使用权限                                                                                                                                                                                  |
|------|------|-------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 规格评估 | MgC  | MgC<br>ServerAssessAge<br>ncyPolicy | ecs:cloudServerFlavors:get ( 查询云服务器规格详情和扩展信息列表 )<br>ims:images:list ( 查询镜像列表 )<br>evs:types:get ( 查询云硬盘类型 )<br>deh:dedicatedHosts:get ( 获取专属主机详情 )<br>deh:dedicatedHosts:list ( 查询专属主机列表 ) |

## 主机绑定目的端

| 场景      | 委托对象 | 自定义策略名称                                 | 细粒度最小使用权限                                                                                                                    |
|---------|------|-----------------------------------------|------------------------------------------------------------------------------------------------------------------------------|
| 主机绑定目的端 | MgC  | MgC<br>ServerBindTarget<br>AgencyPolicy | ecs:cloudServers:showServer ( 查询云服务器详情 )<br>evs:volumes:list ( 查询云硬盘列表 )<br>ecs:cloudServerFlavors:get ( 查询云服务器规格详情和扩展信息列表 ) |

## 主机迁移 workflow

| 场景            | 委托对象 | 自定义策略名称                            | 细粒度最小使用权限                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|---------------|------|------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 主机迁移 workflow | MgC  | MgC<br>ServerMigrationAgencyPolicy | ecs:cloudServers:showServer ( 查询云服务器详情 )<br>ecs:cloudServers:createServers ( 创建云服务器 )<br>sms:server:migrationServer ( 迁移源端服务器 )<br>sms:server:queryServer ( 查看源端服务器 )<br>ecs:cloudServers:list ( 查看云服务器列表 )<br>ecs:cloudServers:listServerBlockDevices ( 查询弹性云服务器磁盘信息 )<br>ecs:cloudServerQuotas:get ( 查询租户配额 )<br>vpc:publicIps:create ( 创建弹性公网IP )<br>vpc:privateIps:list ( 查询私有IP列表 )<br>ecs:cloudServers:get ( 查询云服务器详情 )<br>ecs:cloudServers:changeVpc ( 云服务器切换虚拟私有网络 )<br>ecs:cloudServers:attach ( 弹性云服务器云主机挂载磁盘 )<br>ecs:cloudServers:start ( 批量启动云服务器 )<br>ecs:cloudServers:detachVolume ( 卸载指定弹性云服务器的磁盘 )<br>ecs:cloudServers:stop ( 批量关闭云服务器 )<br>ecs:servers:unlock ( 解锁定云服务器 )<br>evs:volumes:delete ( 删除云硬盘 )<br>evs:volumes:use ( 云硬盘挂载和卸载依赖 )<br>evs:volumes:get ( 查询单个云硬盘详情 )<br>ims:images:get ( 查询镜像详情 ) |

| 场景              | 委托对象 | 自定义策略名称                                  | 细粒度最小使用权限                                                                             |
|-----------------|------|------------------------------------------|---------------------------------------------------------------------------------------|
| 主机迁移工作流开启保持IP不变 | MgC  | MgC<br>ServerMigrationIpKeepAgencyPolicy | vpc:subnets:get ( 查询子网列表或详情 )<br>ecs:cloudServers:listServerInterfaces ( 查询云服务器网卡信息 ) |

## 购买资源

| 场景   | 委托对象 | 自定义策略名称                     | 细粒度最小使用权限                                                                                                                                                                                                                                                                           |
|------|------|-----------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 购买资源 | MgC  | MgC<br>PurchaseAgencyPolicy | eps:resources:add ( 企业项目资源迁入 )<br>ecs:cloudServers:createServers ( 创建云服务器 )<br>evs:volumes:list ( 查询云硬盘列表 )<br>ecs:cloudServerFlavors:get ( 查询云服务器规格详情和扩展信息列表 )<br>ecs:cloudServers:list ( 查询云服务器详情列表 )<br>vpc:publicIps:update ( 更新弹性公网 IP )<br>vpc:publicIps:create ( 创建弹性公网 IP ) |

## 配置主机购买模板

| 场景       | 委托对象 | 自定义策略名称                             | 细粒度最小使用权限                                                                                                                                                                                |
|----------|------|-------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 配置主机购买模板 | MgC  | MgC<br>PurchaseTemplateAgencyPolicy | iam:projects:listProjects ( 查询租户项目 )<br>eps:enterpriseProjects:list ( 查看企业项目列表 )<br>vpc:subnets:get ( 查询子网列表或详情 )<br>vpc:securityGroups:get ( 查询安全组列表或详情 )<br>vpc:vpcs:get ( 查询虚拟私有云详情 ) |

## 迁移方案配置

| 场景                       | 委托对象 | 自定义策略名称                                      | 细粒度最小使用权限                                                                                                                        |
|--------------------------|------|----------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------|
| 主机标准迁移模板 - 导入OBS桶里的目的端配置 | MgC  | MgC<br>ImportTargetConfigurationAgencyPolicy | obs:object:GetObject (获取对象内容、获取对象元数据)<br>obs:bucket:ListBucket (列举桶内对象)<br>obs:bucket:ListAllMyBuckets (获取桶列表)                   |
| 主机标准迁移模板 - 导出支持的目的端配置    |      | MgC<br>ExportTargetConfigurationAgencyPolicy | ims:images:list (查询镜像列表)<br>ecs:cloudServerFlavors:get (查询云服务器规格详情和扩展信息列表)                                                       |
| 主机标准迁移模板 - 主机迁移配额校验      |      | MgC<br>ServerMigrationQuotasAgencyPolicy     | ecs:cloudServers:list (查询云服务器详情列表)<br>evs:quotas:get (查询云硬盘配额)<br>ecs:cloudServerQuotas:get (查询租户配额)<br>vpc:quotas:list (查询资源配额) |
| 对象存储批量迁移模板-配置目的端         |      | MgC<br>ListObsBucketsAgencyPolicy            | obs:bucket:ListBucket (列举桶内对象)<br>obs:bucket:ListAllMyBuckets (获取桶列表)                                                            |

## 创建迁移集群

| 场景     | 委托对象 | 自定义策略名称                         | 细粒度最小使用权限                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|--------|------|---------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 创建迁移集群 | OMS  | OMS<br>ObsMigrationAgencyPolicy | ecs:cloudServers:createServers (创建云服务器)<br>ecs:cloudServers:listServerInterfaces (查询云服务器网卡信息)<br>ecs:cloudServers:showServer (查询云服务器详情)<br>ecs:cloudServers:deleteServers (删除云服务器)<br>nat:natGateways:create (创建NAT网关)<br>nat:natGateways:get (查询NAT网关详情)<br>nat:natGateways:delete (删除NAT网关)<br>nat:snatRules:create (创建SNAT规则)<br>nat:snatRules:get (查询SNAT规则详情)<br>nat:dnatRules:list (查询DNAT规则列表)<br>nat:snatRules:list (查询SNAT规则列表)<br>nat:snatRules:delete (删除SNAT规则)<br>vpc:securityGroups:create (创建安全组)<br>vpc:securityGroups:delete (删除安全组)<br>vpc:securityGroups:get (查询安全组列表或详情)<br>vpc:securityGroupRules:create (创建安全组规则)<br>vpc:securityGroupRules:get (查询安全组规则列表或详情)<br>vpc:securityGroupRules:delete (删除安全组规则)<br>vpcep:epservices:create (创建终端节点服务)<br>vpcep:epservices:get (查询终端节点服务详情)<br>vpcep:epservices:delete (删除终端节点服务) |

| 场景 | 委托对象 | 自定义策略名称                         | 细粒度最小使用权限                                                                                                                                                                                                                                                                 |
|----|------|---------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    |      |                                 | vpcep:connections:update (接受或拒绝终端节点的连接)<br>vpcep:permissions:update (批量添加或删除终端节点服务的白名单)<br>lts:topics:create (创建日志主题)<br>lts:topics:delete (删除指定日志主题)<br>lts:groups:create (创建日志组)<br>lts:groups:get (查询指定日志组)<br>lts:groups:delete (删除指定日志组)<br>ims:images:list (查询镜像列表) |
|    | ECS  | ECS<br>ObsMigrationAgencyPolicy | apm:icmgr:* (apm采集组件的所有权限)                                                                                                                                                                                                                                                |

## 存储迁移 workflow

| 场景            | 委托对象 | 自定义策略名称 | 细粒度最小使用权限                             |
|---------------|------|---------|---------------------------------------|
| 存储迁移 workflow | MgC  | -       | OMS Administrator (系统角色, OMS的admin角色) |

## RVTools 导入

| 场景            | 委托对象 | 自定义策略名称                              | 细粒度最小使用权限                                                                                                      |
|---------------|------|--------------------------------------|----------------------------------------------------------------------------------------------------------------|
| 创建RVTools导入任务 | MgC  | MgC<br>OfflineCollectionAgencyPolicy | obs:object:GetObject (获取对象内容、获取对象元数据)<br>obs:bucket:ListBucket (列举桶内对象)<br>obs:bucket:ListAllMyBuckets (获取桶列表) |

## 1.1.4 IAM 用户权限

MgC在统一身份认证服务（IAM）能力基础上，为用户提供细粒度的权限管理功能，帮助用户灵活便捷地对账号下的IAM用户设置不同的MgC资源权限，结合权限策略实现权限隔离。

以下是使用MgC不同功能时，IAM用户需要配置的权限。为IAM用户配置权限的方法请参考[创建自定义策略](#)。

| 功能            | 权限Action                                                                                                                                                                                                                                                                   |
|---------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 配置主机购买模板      | vpc:vpcs:list（查询虚拟私有云列表）<br>vpc:subnets:get（查询子网列表或详情）<br>vpc:publicips:list（查询弹性公网IP）<br>vpc:securityGroups:get（查询安全组列表或详情）<br>eps:enterpriseProjects:list（查看企业项目列表）<br>ecs:availabilityZones:list（查询可用区列表）                                                               |
| 主机迁移 workflow | vpc:vpcs:list（查询虚拟私有云列表）<br>vpc:vpcs:get（查询虚拟私有云详情）<br>vpc:subnets:get（查询子网列表或详情）<br>vpc:publicips:list（查询弹性公网IP）<br>vpc:publicips:get（查询弹性公网IP详情）<br>vpc:securityGroups:get（查询安全组列表或详情）<br>eps:enterpriseProjects:list（查看企业项目列表）<br>eps:enterpriseProjects:get（查看有企业项目详情） |
| 主机规格评估        | ecs:cloudServerFlavors:get（查询云服务器规格详情和扩展信息列表）<br>ecs:cloudServers:list（查询云服务器详情列表）<br>ecs:cloudServers:showServer（查询云服务器详情）<br>ecs:flavors:get（查询云服务器规格）<br>ims:images:list（查询镜像列表）<br>ims:images:get（查询镜像详情）<br>evs:volumes:list（查询云硬盘列表）<br>evs:types:get（查询云硬盘类型）       |
| AZ迁移 workflow | ecs:availabilityZones:list（查询可用区列表）                                                                                                                                                                                                                                        |
| 配置TCO对比产品映射   | ecs:cloudServerFlavors:get（查询云服务器规格详情和扩展信息列表）<br>ims:images:list（查询镜像列表）<br>evs:types:get（查询云硬盘类型）                                                                                                                                                                         |
| 存储迁移 workflow | Tenant Guest（全部云服务只读权限）（除IAM权限）<br>OMS Administrator（对象存储迁移服务所有权限）                                                                                                                                                                                                         |

| 功能   | 权限Action                                                                                                                                                                                                                                                                                                              |
|------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 迁移集群 | Tenant Guest (全部云服务只读权限) (除IAM权限)<br>OMS Administrator (对象存储迁移服务所有权限)<br>nat:natGateways:list (查询NAT网关列表)<br>smn:topic:list (主题查询权限)                                                                                                                                                                                  |
| 委托授权 | iam:agencies:listAgencies (查询指定条件下的委托列表信息)<br>iam:roles:listRoles (查询权限列表)<br>iam:quotas:listQuotas (查询配额)<br>iam:permissions:listRolesForAgency (查询委托的所有权限)<br>iam:agencies:createAgency (创建委托)<br>iam:permissions:grantRoleToAgency (为委托授予指定权限)<br>iam:roles:createRole (创建自定义策略)<br>iam:roles:updateRole (修改自定义策略) |

## 1.2 配置管理

### 1.2.1 项目管理

#### 功能概述

通过创建独立的项目，可以将不同的操作场景进行区分和隔离，从而实现更好的组织管理和资源分配。包含的项目类型和对应的使用场景说明如下：

- **应用迁移项目**：适用于需要进行主机、存储等资源的采集、迁移场景。
- **复杂迁移（大数据迁移）项目**：适用于需要进行大数据迁移和大数据一致性校验等复杂数据处理场景。

#### 创建项目

- 步骤1** 登录[迁移中心管理控制台](#)。
- 步骤2** 单击左侧导航栏的“**配置管理**”，进入**配置管理**页面。
- 步骤3** 选择“**项目管理**”页签，单击“**创建项目**”，右侧弹出创建新项目窗口。

图 1-2 创建项目



**步骤4** 在弹出的窗口中，仔细查看并了解项目类型的适用场景后，根据实际需求选择项目类型。

### 须知

项目类型选择后将无法更改。

图 1-3 选择项目类型



**步骤5** 输入自定义的项目名称，单击“创建”。创建项目成功，可在项目管理列表查看。

----结束

## 修改项目

**步骤1** 单击操作列的“修改”按钮，弹出“修改项目”窗口。

**步骤2** 可以修改项目名称、输入项目描述、设置项目的开始时间和结束时间，单击“确认”，完成修改。

## 修改项目



\* 项目名称

项目描述

0/255

开始时间

结束时间

确认

取消

----结束

## 归档项目

归档后的项目将在**当前项目**下拉列表中隐藏，无法选择。单击操作列的“**归档**”按钮，可将项目归档。如想重新选择归档后的项目，单击“**取消归档**”即可。



### 说明

当前使用中的项目不支持归档。

## 1.2.2 凭证管理

凭证是资源采集的唯一认证方式，本节介绍如何添加用于发现云上资源（公网发现）所需要的凭证。内网发现、手动添加所需的凭证添加方式请参考[在MgC Agent添加凭证](#)。

## 认证方式

当前仅支持添加公有云资源类型的凭证。支持的认证方式参见[表1-2](#)。

表 1-2 公有云认证方式

| 资源类型 | 认证方式  | 说明                                                                                        |
|------|-------|-------------------------------------------------------------------------------------------|
| 公有云  | AK/SK | 适用于添加 <b>华为云、阿里云、AWS、腾讯云、七牛云和金山云</b> 等云平台凭证。各云平台的AK/SK获取方法请参考 <a href="#">如何获取云平台凭证</a> 。 |

| 资源类型 | 认证方式  | 说明                                                                            |
|------|-------|-------------------------------------------------------------------------------|
|      | 配置文件  | 适用于添加谷歌云平台凭证。配置文件需要上传谷歌云 <b>服务账号</b> 凭据，该凭证必须为 <b>json</b> 格式文件，且大小不能超过4KB。   |
|      | ID/密钥 | 适用于添加 <b>Azure</b> 平台凭证。添加凭证所需的信息获取方法请参见 <a href="#">如何获取添加Azure凭证所需的信息</a> 。 |

**注意**

云上凭证的存储有效期为7天，超过7天后，凭证会过期，需要重新输入。

## 操作步骤

- 步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的**项目**下拉列表中选择**迁移项目**。
- 步骤2** 单击左侧导航栏的“**配置管理**”，进入**配置管理**页面。
- 步骤3** 选择“**凭证管理**”页签，单击“**添加凭证**”，弹出“**添加云平台凭证**”窗口。

图 1-4 添加凭证



- 步骤4** 输入自定义的凭证名称，选择认证方式并输入/上传相应凭证，单击“**确认**”。添加完成后，在凭证管理列表可以看到添加的凭证。

图 1-5 选择认证方式



----结束

## 1.3 调研评估

### 1.3.1 IDC 上云成本对比

通过输入IDC的详细信息和相关假设条件，系统将自动匹配源端IDC与华为云的对应产品及用量，并计算费用对比，为您提供上云迁移的参考依据和相关云上使用成本的分析，方便您做出上云决策。

#### 创建任务

- 步骤1** 登录[迁移中心](#)管理控制台。在左侧导航栏的**项目**下拉列表中选择**迁移项目**。
- 步骤2** 单击左侧导航栏“**调研评估>上云调研**”，进入“**上云调研**”页面。
- 步骤3** 选中**IDC上云成本对比**框，单击“**创建任务**”按钮，进入“**IDC上云成本对比**”基础配置页面。



- 步骤4** 设置“**数据中心**”区域参数，具体参数参见[表1-3](#)。

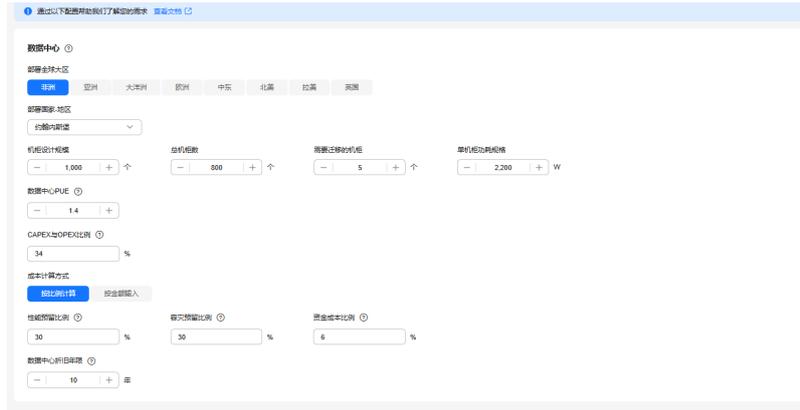


表 1-3 设置数据中心说明参数

| 参数           | 说明                                                                                                  |
|--------------|-----------------------------------------------------------------------------------------------------|
| 部署全球大区       | 选择部署全球大区，包括：非洲，亚洲，大洋洲，欧洲，中东，北美，拉美，英国。                                                               |
| 部署国家-地区      | 下拉选择，根据选择部署全球大区而出现不同选项。                                                                             |
| 机柜设计规模       | 默认值1000个， $1 \leq \text{取值范围} \leq 99,999$ ，取整数。<br><b>说明</b><br>机柜设计规模 $\geq$ 总机柜数 $\geq$ 需要迁移的机柜。 |
| 总机柜数         | 默认值800个， $1 \leq \text{取值范围} \leq 99,999$ ，取整数。<br><b>说明</b><br>机柜设计规模 $\geq$ 总机柜数 $\geq$ 需要迁移的机柜。  |
| 需要迁移的机柜      | 默认值5个， $1 \leq \text{取值范围} \leq 99,999$ ，取整数。<br><b>说明</b><br>机柜设计规模 $\geq$ 总机柜数 $\geq$ 需要迁移的机柜。    |
| 单机柜功耗规格      | 默认值2200w， $1000 \leq \text{取值范围} \leq 30,000$ ，取整数。                                                 |
| 数据中心PUE      | 评价数据中心能源效率的指标，为数据中心能耗的所有能源与IT负载使用的能源之比。<br>默认值1.4， $1 < \text{取值范围} < 10$ 。                         |
| CAPEX与OPEX比例 | 数据中心投资成本与运营成本的比例。<br>默认值34%， $1 < \text{取值范围} < 1000$ 。                                             |

| 参数       | 说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 成本计算方式   | <p>根据需求选择不同的成本计算方式，以有效实现有效成本管理和控制。</p> <ul style="list-style-type: none"> <li>选择按比例计算，需要设置以下参数： <ul style="list-style-type: none"> <li>性能预留比例：默认值30%，1≤取值范围≤100。</li> <li>容灾预留比例：默认值30%，1≤取值范围≤100。</li> <li>资金成本比例：默认值6%，1≤取值范围≤100。</li> </ul> </li> <li>选择按金额输入，需要设置以下参数： <ul style="list-style-type: none"> <li>性能预留成本：默认值100000\$，1≤取值范围≤10000000000000000。</li> <li>容灾预留成本：默认值10000\$，1≤取值范围≤10000000000000000。</li> <li>资金成本：默认值10000\$，1≤取值范围≤10000000000000000。</li> </ul> </li> </ul> |
| 数据中心折旧年限 | 默认值10，1≤取值范围≤20，取整数。                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |

**步骤5** 设置“服务器”区域参数，具体参数说明参见[表1-4](#)。

### 服务器

**CPU总核数** ?

-
1,000
+
个

**高性能CPU比例** ?

%

**高性价比CPU比例** ?

%

**成本优先CPU比例** ?

%

**是否涉及Vmware虚拟化**

是
否

表 1-4 设置服务器参数

| 参数            | 说明                                                                                                                                                                                      |
|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| CPU总核数        | 环境中需要迁移的服务器CPU数量。默认值1000个，1≤取值范围≤99,999,999。                                                                                                                                            |
| 高性能CPU比例      | 如数据库、流媒体、消息服务器、推理等(C3)，华为云对应规格：C7等，默认值为20%。<br><b>说明</b><br>持续高性能、成本和性能平衡、成本优先的比例相加为100%。                                                                                               |
| 高性能比CPU比例     | 如微服务、Web和应用服务器、BI等(N2)，华为云对应规格：aC7等，默认值为20%。<br><b>说明</b><br>持续高性能、成本和性能平衡、成本优先的比例相加为100%。                                                                                              |
| 成本优先CPU比例     | 如开发/测试、内部应用程序、批处理等(E2)，华为云对应规格：S7，HECS等，默认值为60%。<br><b>说明</b><br>持续高性能、成本和性能平衡、成本优先的比例相加为100%。                                                                                          |
| 是否涉及Vmware虚拟化 | 计算物理核心上托管的虚拟CPU的所有数量。如果工作负载包含虚拟机，则使用此数据来计算底层物理服务器的数量。<br><ul style="list-style-type: none"> <li>选择“否”，不涉及Vmware虚拟化。</li> <li>选择“是”，设置虚拟化的vCPU数量。默认值300个，1≤取值范围&lt;100000000。</li> </ul> |

步骤6 设置“存储”区域参数，具体参数说明参见表1-5。



表 1-5 设置存储参数

| 参数      | 子参数  | 说明                                                                                                                                                                            |
|---------|------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 迁移总存储容量 | -    | 默认值36，0≤取值范围≤999,999。单位为TB。                                                                                                                                                   |
| 存储类型占比  | 块存储  | 比例：设置块存储占总存储容量的比例。例如：总存储容量为100 TB，块存储比例设置为20%，则块存储容量为20 TB。<br><ul style="list-style-type: none"> <li>HDD（普通硬盘）：设置HDD占块存储容量的比例。</li> <li>SSD（固态硬盘）：设置SSD占块存储容量的比例。</li> </ul> |
|         | 文件存储 | 比例：设置文件存储占总存储容量的比例。例如：总存储容量为100 TB，文件存储比例设置为60%，则文件存储容量为60 TB。                                                                                                                |

| 参数 | 子参数  | 说明                                                                                                                                                                                                                                                                                                    |
|----|------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    | 对象存储 | <p>比例：设置对象存储占总存储容量的比例。例如：总存储容量为100 TB，对象存储比例设置为20%，则对象存储容量为20 TB。</p> <ul style="list-style-type: none"> <li>标准桶：设置标准桶占对象存储容量的比例。</li> <li>低频桶：设置低频桶占对象存储容量的比例。</li> <li>归档桶：设置归档桶占对象存储容量的比例。</li> <li>深度归档桶：设置深度归档桶占对象存储容量的比例。</li> </ul> <p><b>说明</b><br/>存储桶类型的详细说明请参考：<a href="#">存储类别简介</a>。</p> |

**步骤7** 设置“网络”区域参数，月度网络带宽量通过拉取动态条或手工直接填写，系统会根据默认价格进行月单价计算，也可以根据您的实际情况进行调整。



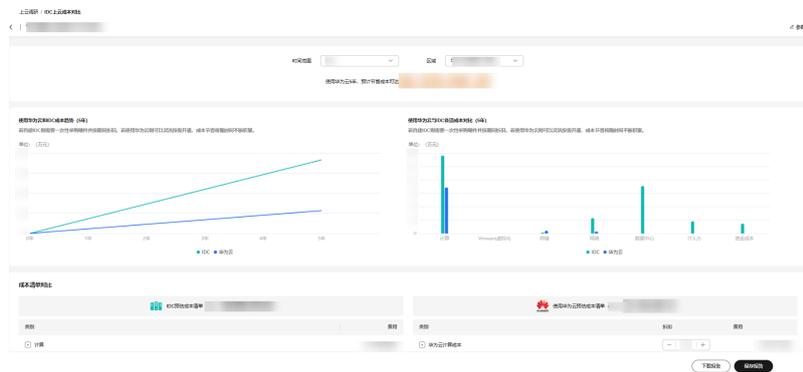
**步骤8** 单击下一步，进入“调整假定条件”页面，设置对应参数，具体参数说明参见表1-6。

**表 1-6** 设置调整假定条件页面参数

| 区域     | 参数                | 说明                                                     |
|--------|-------------------|--------------------------------------------------------|
| 数据中心成本 | 计算周期              | 1≤取值范围≤5，取整数。                                          |
|        | 数据中心单瓦（W）建设参考成本   | 基于基础配置页面选择的数据中心-部署全球大区及部署国家-地区，确定参考建设成本。也可以根据实际情况自行填写。 |
| 硬件成本   | 云下服务器成本/核/年       | 取物理机价格除以核数，求平均值。默认值为市场参考价，可根据实际情况设置。                   |
| 软件成本   | 服务器软件成本比例/年       | 设置服务器软件成本占全年软件成本的比例，根据实际情况设置。                          |
|        | 存储软件成本比例/年        | 设置存储软件成本占全年软件成本的比例，根据实际情况设置。                           |
| 虚拟化成本  | VMware虚拟化软件成本/核/年 | 每个物理服务器上运行的VMware虚拟化软件成本，可根据实际情况设置。                    |
| 网络成本   | 网络成本占服务器成本百分比     | 设置网络成本占服务器成本的比例，根据实际情况设置。                              |
|        | 服务提供商成本/月         | 基于基础配置页面设置的月度网络带宽量，确定每月服务提供商成本。也可以根据实际情况自行填写。          |

| 区域     | 参数                 | 说明                                                                                                                                                                                                                                       |
|--------|--------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 存储成本   | 块存储                | <ul style="list-style-type: none"> <li>HDD成本/GB/月：设置每GB HDD所花费的成本，可根据实际情况设置。</li> <li>SSD成本/GB/月：设置每GB SSD所花费的成本，可根据实际情况设置。</li> </ul>                                                                                                   |
|        | 文件存储               | 文件存储成本/GB/月：设置每GB 文件存储所花费的成本，可根据实际情况设置。                                                                                                                                                                                                  |
|        | 对象存储               | <ul style="list-style-type: none"> <li>标准桶成本/GB/月：设置每GB 标准桶所花费的成本，可根据实际情况设置。</li> <li>低频桶成本/GB/月：设置每GB 低频桶所花费的成本，可根据实际情况设置。</li> <li>归档桶成本/GB/月：设置每GB 归档桶所花费的成本，可根据实际情况设置。</li> <li>深度归档桶成本/GB/月：设置每GB 深度归档桶所花费的成本，可根据实际情况设置。</li> </ul> |
| IT人力成本 | 硬件维护成本占硬件成本比例      | 根据实际情况设置。                                                                                                                                                                                                                                |
|        | 网络维护成本占网络硬件和软件成本比例 |                                                                                                                                                                                                                                          |
|        | 存储维护成本占存储采购成本比例    |                                                                                                                                                                                                                                          |

**步骤9** 单击“查看报告”，进入报告预览页面。报告预览页面根据选择时间范围和区域可以查看：预计节省成本数值，使用华为云和IDC成本趋势、各项成本对比，成本清单对比等详细信息。



报告预览页面可进行如下操作：

- **调整时间范围和区域**：不同时间范围和不同区域预计节省成本也会不一样；
- **调整成本清单对比折扣**：根据从华为云商务代表处获得的实际折扣，调整各类别的优惠折扣，系统会根据设置的折扣计算出华为云预估成本清单。

- **参数调整**：调整成本任务参数配置。

**步骤10** 选择报告的保存方式：

- 单击“**下载报告**”，将报告下载到本地查看。
- 单击“**保存报告**”，将报告保存在报告列表中，单击报告名称可以在线查看和重新**调整报告**。



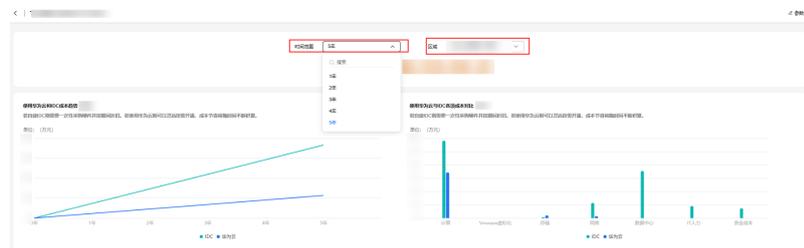
----结束

## 时间范围和区域

您可以在成本对比参数设置完成后，查看不同时间范围和区域下所包含的预计节省成本数值，使用华为云和IDC成本趋势、各项成本对比，成本清单对比等详细信息。

在IDC上云成本对比任务列表，单击**报告名称**，进入报告详情页面。在报告详情页面，即可设置时间范围和区域。

图 1-6 时间范围和区域



## 成本清单对比折扣

您可以通过和**华为云商务代表**沟通，针对部分类别产品分别调整优惠折扣，进一步优化**华为云**预估成本清单。

在IDC上云成本对比任务列表，单击**报告名称**，进入报告详情页面。在成本清单对比模块，可调整折扣。

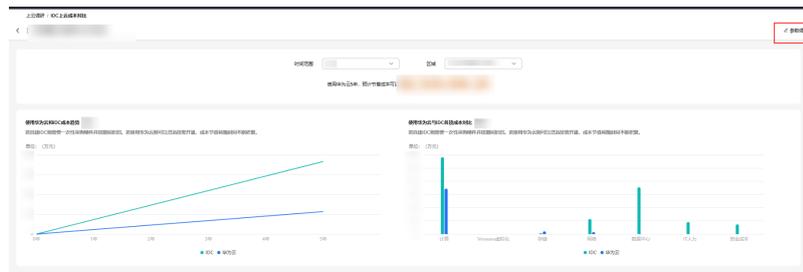
图 1-7 成本清单对比

## 参数调整

您可以在查看报告后，对参数需要调整的地方进行参数重新配置，重新生成报告。

在IDC上云成本对比任务列表，单击**报告名称**，进入报告详情页面。在报告右上角单击“**参数调整**”。

图 1-8 参数调整



## 调整报告

**步骤1** 在报告列表中，单击报告名称可以在线查看和重新编辑报告内参数。

**步骤2** 单击保存报告，选择保存方式：

- 保存为新报告：修改报告名称，单击确认，报告列表中新增该报告。
- 覆盖原有报告：单击确认，报告列表中原有报告根据修改内容刷新。



----结束

### 1.3.2 跨云迁移成本分析（基于凭证）

通过输入源端云平台凭证，可以自动完成源端云厂商消费账单分析，根据源端产品规格及用量自动匹配华为云上的产品及用量，并对比源端云厂商使用费用和华为云使用费用，为您云间迁移提供参考。

目前支持通过凭证进行上云成本分析的源端云厂商包括：**AWS**、**阿里云**。

## 须知

进行上云成本分析时，请确保服务器数量不超过9999台。

## 创建分析任务

- 步骤1** 登录**迁移中心**管理控制台。在左侧导航栏的**项目**下拉列表中选择**迁移项目**。
- 步骤2** 单击左侧导航栏“**调研评估>上云调研**”，进入“**上云调研**”页面。
- 步骤3** 选中**跨云迁移成本分析框**，在上云分析页签，单击“**创建分析任务**”按钮，首次使用该功能时需要授权委托，同意授权后右侧弹出“**创建分析任务**”窗口。

图 1-9 创建上云成本分析任务



- 步骤4** 根据**表1-7**，配置参数。

**创建分析任务**

**任务**

任务名称

**源端**

云厂商

区域

**调研方式**

AWS/CK  表格上传

**凭证**

请选择凭证

**对比时间段**

开始月份—结束月份

**对比产品**

请选择产品

**目的端**

目的端区域

表 1-7 参数说明

| 参数   | 说明                                          |
|------|---------------------------------------------|
| 任务名称 | 用户自定义。                                      |
| 云厂商  | 选择源端云厂商。<br><b>说明</b><br>当前源端云厂商仅支持AWS和阿里云。 |
| 区域   | 选择源端资源所在区域。                                 |

| 参数    | 说明                                                                                                                                                                                                                                                                                                                                                                                |
|-------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 调研方式  | <ul style="list-style-type: none"> <li>“AK/SK”方式，需要选择“凭证”、“对比时间段”和“对比产品”，如未提前<a href="#">添加凭证</a>，请单击“<a href="#">新建凭证</a>”，填写相应参数，校验并保存凭证。<br/><b>说明</b> <ul style="list-style-type: none"> <li>新建凭证中的AK/SK输入源端账号的AK/SK。</li> <li>源端账号拥有只读权限即可。</li> <li>可能产生接口调用相关费用。</li> </ul> </li> <li>“表格上传”方式，通过填写TCO任务模板的方式上传源端信息进行评估分析。<br/><b>须知</b><br/>模板中的主机数量不能超过9999台。</li> </ul> |
| 目的端区域 | 选择目的端所在区域。                                                                                                                                                                                                                                                                                                                                                                        |

**步骤5** 配置完成后，单击“[确认](#)”按钮，提交上云成本分析任务。

提交成功后可以在上云分析任务列表中看到提交的任务，单击[任务名称](#)，可以查看任务进度，等待任务运行完成后即可[下载账单](#)、[查看配置](#)、[产品对比](#)，进行后续详细分析。

----结束

## 下载账单

系统会从产品计费项的维度分析您指定时间段内的账单，解析出每款产品的规格、月均使用量以及月均消耗，您可以在上云成本分析任务完成后下载您的账单，在本地进一步分析。

**步骤1** 在上云分析任务列表，单击[任务名称](#)。

**步骤2** 单击操作列的“[下载账单](#)”按钮，下载Excel版账单。



----结束

## 查看配置

系统会根据解析出来的产品规格及用量自动匹配华为云上的产品及用量，在匹配策略上系统坚持在不低于您当前配置的前提下优先考虑成本的原则，给出华为云上的产品规格及用量配置。

您可以通过[查看配置](#)进入产品映射明细列表，在这个列表中，您可以添加、修改、删除产品和规格映射，调整程序预定义的资源映射逻辑，从而获取您真实需要的产品规格列表。

**步骤1** 在上云分析任务列表，单击[任务名称](#)。

**步骤2** 单击操作列的“查看配置”按钮，进入“产品映射”页面。



**步骤3** 单击映射类型前的按钮，可以查看映射详情。

#### 说明

- 映射：指在不同产品或规格之间建立的对应关系。
- 标准映射：指迁移中心推荐的源端产品规格到华为云的对应产品规格。
- 非标准映射：指在没有标准映射产品规格可用时，迁移中心选择的最接近的替代产品规格，或者用户根据需求手动更改后所选择的产品规格。

---结束

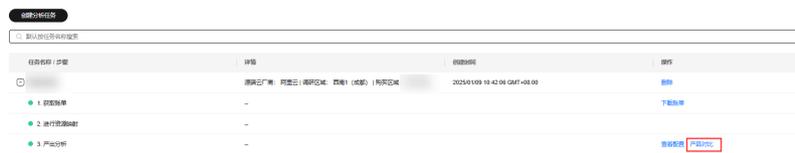
## 产品对比

在使用成本方面，系统会对指定时间内，源端云厂商使用费用和华为云使用费用进行比较，您可以在**产品对比**页面的**对比结果**中查看。

我们还会给出源端云厂商和华为云同等规格配置的产品月均成本比较，您可以在**产品对比**页面的**产品价格对比**中查看。并且可以通过和您的销售经理沟通，针对每种类别产品分别调整价格计算比例（优惠折扣）进一步优化华为云的使用成本。

**步骤1** 在上云分析任务列表，单击**任务名称**。

**步骤2** 单击操作列的“产品对比”按钮，进入“产品对比”页面。



**步骤3** 查看**成本对比结果**和**产品价格对比**，根据实际情况调整价格计算比例（优惠折扣）优化华为云的使用成本。

单击页面右上角“导出结果”按钮，可以将产品对比结果导出PDF格式到本地。

---结束

### 1.3.3 跨云迁移成本分析（基于账单）

通过上传源端云厂商账单导出件，可以自动完成源端云厂商消费账单分析，根据源端产品规格及用量自动匹配华为云上的产品及用量，并对比源端云厂商使用费用和华为云使用费用，为您云间迁移提供参考。

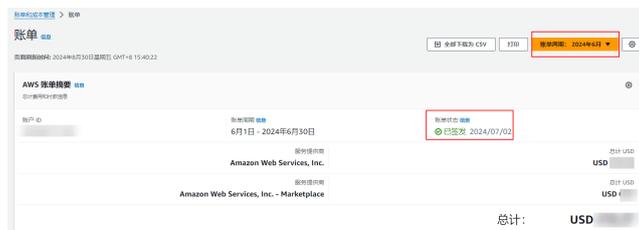
## 约束与限制

- 目前支持通过账单进行上云成本分析的源端云厂商有：**AWS**。
- 当前支持 EC2/RDS/S3 等17种产品的自动映射，不支持的产品将不会进行产品映射和统计。

- Amazon EC2 (Compute)
  - Amazon EC2 (EBS)
  - Amazon EC2 (NAT Gateway)
  - Amazon RDS (MySQL)
  - Amazon RDS (AuroraMySQL)
  - Amazon EKS (Fargate)
  - Amazon ECS (Fargate)
  - Amazon DataTransfer
  - Amazon VPC (VPN)
  - Amazon VPC (VPCEP)
  - Amazon ElastiCache
  - Amazon DocDB
  - Amazon ELB
  - Amazon S3
  - Amazon DynamoDB
  - Amazon CloudWatch
  - Amazon ECR
- 每个MgC项目下，最多创建100个分析任务，且任务不能重名。
  - 上传的账单文件必须满足如下要求：
    - 格式必须为英语或葡萄牙语的文字类型PDF，不支持图片类型的PDF。
    - 文件大小不超过50 MB。
    - 文件名中不能包含空格。

## 准备工作

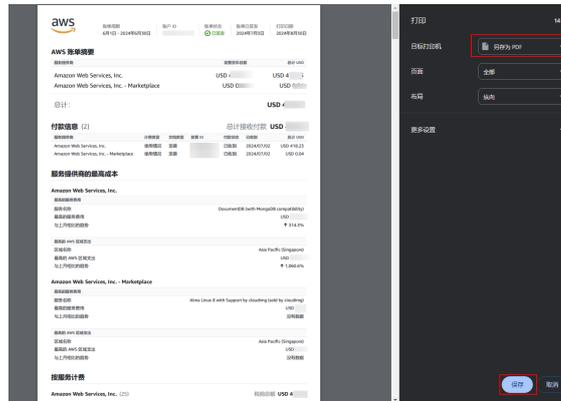
- **创建项目：**  
在MgC控制台创建独立的项目，方法请参考[项目管理](#)。
- **下载AWS账单**
  - a. 打开 **AWS Billing and Cost Management 控制台**。
  - b. 在账单页面，选择需要下载账单的月份。在 AWS 账单摘要，确认账单状态显示为“已签发”。



- c. 单击页面右上角的“打印”按钮，勾选需要打印的内容，其中**最高成本**、**按服务计费**为必选项，单击“打印”。



d. 在弹出的打印窗口，选择另存为PDF，单击“保存”。



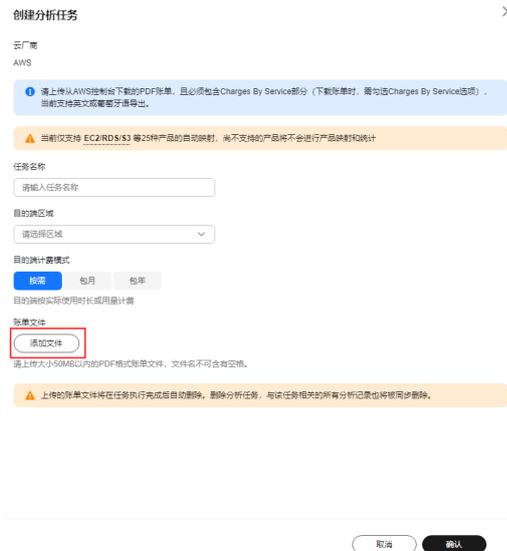
## 创建分析任务

- 步骤1** 登录**迁移中心管理控制台**。在左侧导航栏的**项目**下拉列表中选择**迁移项目**。
- 步骤2** 单击左侧导航栏“**调研评估>上云调研**”，进入“**上云调研**”页面。
- 步骤3** 选中**跨云迁移成本分析框**，在**账单导入**页签，单击“**创建分析任务**”按钮，首次使用该功能时需要授权委托，同意授权后右侧弹出“**创建分析任务**”窗口。

图 1-10 创建账单分析任务



- 步骤4** 自定义任务名称，并选择目的端所在区域后，单击“**添加文件**”，将导出的账单文件上传至MgC。



**步骤5** 账单上传成功后，单击“**确认**”按钮，提交账单导入成本分析任务。提交成功后可以在账单导入任务列表中查看任务状态、源端总成本和华为云总成本等详细信息。

当任务状态为“**已完成**”时，可以进行如下操作：

- **查看产品映射详情**：可以查看账单内包含的产品类型、数量以及每个产品的映射状态和结果。
- **修改映射结果**：支持手动调整产品映射结果。
- **调整价格计算比例**：通过调整华为云产品的优惠折扣，进一步优化华为云的使用成本。
- **导出成本分析结果**：将分析结果明细下载到本地进行手动调整和分析。

----**结束**

## 查看产品映射详情

您可以在成本分析任务完成后，查看账单内包含的所有产品类型、数量以及每个产品的映射状态和结果。

在账单导入任务列表，单击**任务名称**，进入任务详情页面。在任务详情页面，即可查看产品映射详情。

**图 1-11** 产品映射详情



## 修改映射结果

对于映射成功或失败的产品，均支持手动修改目的端产品类型、目的端产品和规格、目的端使用量等参数，从而调整目的端月度成本。

**步骤1** 在账单导入任务列表，单击**任务名称**，进入任务详情页面。

**步骤2** 在产品映射列表，单击操作列的“**修改**”，右侧弹出修改窗口。

图 1-12 修改映射结果

| 源端产品        | 源端产品        | 计费周期        | 数量        | 映射状态 | 源端产品      | 目的端产品      | 使用量  | 计费模式 | 源端成本    | 目的端成本 | 折扣      | 操作    |
|-------------|-------------|-------------|-----------|------|-----------|------------|------|------|---------|-------|---------|-------|
| Data Tran.  | Data Tran.  | Asia Pacif. | 1,879,305 | 成功映射 | Cloud Co. | Global Co. | --   | 按量   | \$ 0.00 | 0     | \$ 0.00 | 修改 删除 |
| Simple St.  | Simple St.  | Asia Pacif. | 374,133.3 | 成功映射 | Cloud St. | Cloud St.  | size | 按量   | \$ 0.00 | 374   | \$ 0.00 | 修改 删除 |
| Elastic Co. | Elastic Co. | Asia Pacif. | 10,079.71 | 映射失败 | --        | --         | --   | 按量   | \$ 0.00 | 0     | \$ 0.00 | 修改 删除 |
| Data Tran.  | Data Tran.  | Asia Pacif. | 45,587.82 | 成功映射 | --        | --         | --   | 按量   | \$ 0.00 | 0     | \$ 0.00 | 修改 删除 |

**步骤3** 在目的端产品信息区域，根据实际需求，选择目的端产品类型、目的端产品和规格，系统会根据您的修改，重新计算目的端月度成本。

确认调整目的端产品类型和规格后，单击“**保存**”按钮。

图 1-13 修改目的端规格

目的端产品信息

目的端区域

计费模式

按量

目的端产品类型

对象存储服务

目的端产品

云存储

规格选择: 对象存储服务(云存储) direct access plat(GB)

| 目的端产品类型                          | 目的端产品  | 资源规格 | 用量类型 (单位) |    |
|----------------------------------|--------|------|-----------|----|
| <input checked="" type="radio"/> | 对象存储服务 | 云存储  | 混合平台      | GB |
| <input type="radio"/>            | 对象存储服务 | 云存储  | 固睿平台      | GB |

目的端使用量

3 GB

目的端月度成本

取消 保存

----结束

## 调整价格计算比例

您可以通过和为云商务代表沟通，针对每种类别产品分别调整价格计算比例（优惠折扣），进一步优化华为云的使用成本。

**步骤1** 在账单导入任务列表，单击**任务名称**，进入任务详情页面。

**步骤2** 在页面右上角，单击折扣后的“**修改**”，右侧弹出折扣设置窗口。

图 1-14 修改折扣比例



**步骤3** 根据从华为云商务代表处获得的实际折扣，调整各产品的优惠折扣。折扣设置包括两部分：总折扣和分类产品折扣。

- 总折扣：修改总折扣会同步修改所有产品的折扣。您可以先调整总折扣，然后根据需要单独调整每个分类产品的折扣。
- 分类产品折扣设置：根据不同类型产品的实际折扣设置，修改分类产品折扣后不会同步修改总折扣。在计算最终折扣价格时，将使用分类产品的折扣。

**步骤4** 所有产品折扣设置完成后，单击“**保存**”按钮。系统会根据设置的折扣计算出华为云最终的使用成本。

----结束

## 导出成本分析结果

您可以在成本分析任务完成后，将分析结果明细下载到本地进行手动调整和分析。

- **方法一**：在账单导入任务列表，单击任务操作列的“**导出**”，下载该任务的分析结果明细（Excel文件）。
- **方法二**：在任务详情页面，单击页面右上角的“**导出**”按钮，下载该任务的分析结果明细（Excel文件）。



## 1.3.4 问卷调研

### 概述

通过问卷调研的方式，基于华为云专业服务多年的经验积累，为您提供全面的源端架构梳理及资源分析功能。

### 运维架构调研

**步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的**项目**下拉列表中选择**迁移项目**。

**步骤2** 单击左侧导航栏“**调研 > 上云调研**”，进入“**上云调研**”页面。

**步骤3** 选中**问卷调研**框，单击“**运维架构调研**”框中的“**开始**”按钮，进入运维架构调研问答页面。

图 1-15 开始调研

#### 问卷调研



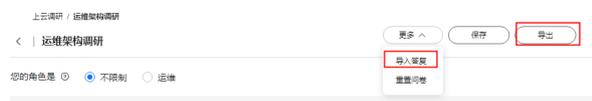
**步骤4** 根据实际情况，选择您的角色。选择角色仅用于筛选问题，不会影响文件内容。

**图 1-16 选择角色**



**步骤5** 依次填写调研问卷所包含的问题。可以参考页面右侧的回答样例。如果在填写问卷时有疑问，或者需要咨询专业团队的意见，可以单击页面右上角的“导出”按钮，导出问卷，待专业团队进行分析并答复后，再导入问卷。

**图 1-17 导出、导入问卷**



**步骤6** 确定所有问卷问题的答复后，单击页面右上角的“保存”按钮，运维架构调研完成。

----结束

### 1.3.5 存储迁移成本调研

迁移中心 MgC提供了存储迁移成本评估功能，通过输入源端文件数量和大小、设定期望迁移时长等参数，系统会自动计算迁移所需的集群规格、网络带宽和迁移成本，并给出不同的迁移方案，用户可以根据项目需求和预算选择合适的迁移策略。

#### 操作步骤

**步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的**项目**下拉列表中选择**迁移项目**。

**步骤2** 单击左侧导航栏“**调研评估>上云调研**”，进入“**上云调研**”页面。

**步骤3** 单击**存储迁移成本调研**框，进入存储迁移成本调研参数配置页面。

**步骤4** 在**源端详情**区域，根据实际情况，选择源端平台和源端类型，并输入源端文件个数、总文件大小。

**步骤5** 在**目的端详情**区域，选择目的端区域和目的端类型。

**步骤6** 在**高级特性**选择网络模式、偏好设置，并输入期望迁移时间。

**步骤7** 配置完以上参数后，单击“**创建评估**”按钮。系统会根据配置参数，并给出不同的迁移方案。

----结束

## 1.4 资源采集

## 1.4.1 在线调研

通过公网在线发现并采集源端云平台资源，获取源端资源详细信息，并整理源端资源与应用间的关联关系。目前支持发现的源端云平台为：**阿里云、华为云、AWS、腾讯云、谷歌云、Azure、七牛云、金山云**。

- **阿里云和华为云**支持采集的资源类型包括：**主机、容器、中间件、数据库、网络和存储**。
- **腾讯云**支持采集的资源类型包括：**主机、数据库、存储**。
- **谷歌云、AWS和Azure**支持采集的资源类型包括：**主机、容器、中间件、数据库、存储和网络**。
- **七牛云和金山云**支持采集的资料类型包括：**对象存储**。

### 创建在线调研任务

- 步骤1** 登录**迁移中心管理控制台**。在左侧导航栏的项目下拉列表中选择创建的**应用迁移项目**。
- 步骤2** 单击左侧导航栏的“**源端发现>资源采集**”，进入**资源采集**页面。
- 步骤3** 选择“**在线调研**”页签，单击“**云平台采集**”，进入**云平台采集**参数配置页面。

图 1-18 云平台采集



- 步骤4** 根据**表1-8**，配置采集任务参数。

表 1-8 云平台采集参数说明

| 区域   | 参数   | 说明                                                           | 是否必选 |
|------|------|--------------------------------------------------------------|------|
| 基础配置 | 任务名称 | 自定义发现任务名称。                                                   | 是    |
|      | 任务说明 | 输入发现任务说明。                                                    | 否    |
| 任务配置 | 源端   | 选择源端云平台，目前支持的云平台有 <b>阿里云、华为云、AWS、腾讯云、谷歌云、Azure、七牛云、金山云</b> 。 | 是    |

| 区域 | 参数   | 说明                                                                                                                                                                                                                                                                                                                                                                  | 是否必选 |
|----|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
|    | 凭证   | <p>选择源端对应的凭证，如未提前<a href="#">添加凭证</a>，请单击“<a href="#">新建凭证</a>”，填写相应参数，校验并保存凭证。</p> <ul style="list-style-type: none"> <li>• 阿里云、华为云、AWS、腾讯云、七牛云、金山云认证方式请选择“AK/SK”，并输入源端账号的AK/SK。</li> <li>• 谷歌云认证方式请选择“<a href="#">配置文件</a>”，并上传谷歌云服务账号凭据，该凭证必须为json格式文件，且大小不能超过4KB。</li> <li>• Azure认证方式请选择“ID/密钥”。添加凭证所需的信息获取方法请参见<a href="#">如何获取添加Azure凭证所需的信息</a>。</li> </ul> | 是    |
|    | 区域   | 选择要发现的源端云平台区域，支持多选。                                                                                                                                                                                                                                                                                                                                                 | 是    |
|    | 资源类型 | 在下拉列表，勾选需要进行采集的资源类型（采集项）。当前支持采集的云平台和资源类型参见 <a href="#">表1-9</a> 。                                                                                                                                                                                                                                                                                                   | 是    |
|    | 应用   | <p>选择采集到的资源要加入的应用，通过应用可以将资源进行分组管理。如果下拉列表中无应用，按如下步骤创建：</p> <ol style="list-style-type: none"> <li>1. 单击“<a href="#">新建应用</a>”，输入自定义的<a href="#">应用名称</a>和<a href="#">描述</a>；选择<a href="#">业务场景</a>和<a href="#">使用场景</a>；目的端区域选择目的端所在区域。</li> <li>2. 单击“<a href="#">确定</a>”按钮，完成创建。</li> </ol>                                                                        | 否    |

表 1-9 支持采集的资源类型

| 云平台 | 资源类型                                                                                               | 子类型 |
|-----|----------------------------------------------------------------------------------------------------|-----|
| 阿里云 | <ul style="list-style-type: none"> <li>• 主机</li> <li>• 容器</li> <li>• 大数据</li> <li>• 数据库</li> </ul> | -   |

| 云平台 | 资源类型                                                                                               | 子类型                                                                                                                                                                                                                                 |
|-----|----------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|     | 中间件                                                                                                | <ul style="list-style-type: none"> <li>• Redis</li> <li>• Kafka</li> </ul>                                                                                                                                                          |
|     | 存储                                                                                                 | <ul style="list-style-type: none"> <li>• 对象存储</li> <li>• 文件存储</li> </ul>                                                                                                                                                            |
|     | 网络                                                                                                 | <ul style="list-style-type: none"> <li>• 云连接</li> <li>• 负载均衡（ALB、CLB）</li> <li>• 专线</li> <li>• 公网域名</li> <li>• 内网域名</li> <li>• 弹性公网IP</li> <li>• 公网NAT网关</li> <li>• 路由表</li> <li>• 安全组</li> <li>• 虚拟私有云</li> <li>• VPN网关</li> </ul> |
| 华为云 | <ul style="list-style-type: none"> <li>• 主机</li> <li>• 容器</li> <li>• 大数据</li> <li>• 数据库</li> </ul> | -                                                                                                                                                                                                                                   |
|     | 中间件                                                                                                | <ul style="list-style-type: none"> <li>• Redis</li> <li>• Kafka</li> </ul>                                                                                                                                                          |
|     | 存储                                                                                                 | <ul style="list-style-type: none"> <li>• 对象存储</li> <li>• 文件存储</li> </ul>                                                                                                                                                            |
|     | 网络                                                                                                 | <ul style="list-style-type: none"> <li>• 负载均衡（ELB）</li> <li>• 公网域名</li> <li>• 内网域名</li> <li>• 弹性公网IP</li> <li>• 公网NAT网关</li> <li>• 路由表</li> <li>• 安全组</li> <li>• 虚拟私有云</li> </ul>                                                   |
| AWS | <ul style="list-style-type: none"> <li>• 主机</li> <li>• 容器</li> <li>• 数据库</li> </ul>                | -                                                                                                                                                                                                                                   |

| 云平台   | 资源类型                                                                          | 子类型                                                                                                                                                                           |
|-------|-------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|       | 中间件                                                                           | <ul style="list-style-type: none"> <li>Redis</li> <li>Kafka</li> </ul>                                                                                                        |
|       | 网络                                                                            | <ul style="list-style-type: none"> <li>负载均衡（ELB）（目前只支持CLB类型）</li> <li>公网域名</li> <li>内网域名</li> <li>弹性公网IP</li> <li>公网NAT网关</li> <li>路由表</li> <li>安全组</li> <li>虚拟私有云</li> </ul> |
|       | 存储                                                                            | <ul style="list-style-type: none"> <li>对象存储</li> <li>文件存储</li> </ul>                                                                                                          |
| 腾讯云   | <ul style="list-style-type: none"> <li>主机</li> <li>数据库</li> </ul>             | -                                                                                                                                                                             |
|       | 存储                                                                            | <ul style="list-style-type: none"> <li>对象存储</li> <li>文件存储</li> </ul>                                                                                                          |
| 谷歌云   | <ul style="list-style-type: none"> <li>主机</li> <li>容器</li> <li>数据库</li> </ul> | -                                                                                                                                                                             |
|       | 中间件                                                                           | Redis                                                                                                                                                                         |
|       | 存储                                                                            | <ul style="list-style-type: none"> <li>对象存储</li> <li>文件存储</li> </ul>                                                                                                          |
|       | 网络                                                                            | <ul style="list-style-type: none"> <li>弹性公网IP</li> <li>路由表</li> <li>安全组</li> <li>虚拟私有云</li> </ul>                                                                             |
| Azure | <ul style="list-style-type: none"> <li>主机</li> <li>容器</li> <li>数据库</li> </ul> | -                                                                                                                                                                             |
|       | 存储                                                                            | <ul style="list-style-type: none"> <li>对象存储</li> <li>文件存储</li> </ul>                                                                                                          |
|       | 中间件                                                                           | <ul style="list-style-type: none"> <li>Redis</li> <li>Kafka</li> </ul>                                                                                                        |

| 云平台 | 资源类型 | 子类型                                                                                                                                   |
|-----|------|---------------------------------------------------------------------------------------------------------------------------------------|
|     | 网络   | <ul style="list-style-type: none"> <li>弹性公网IP</li> <li>路由表</li> <li>安全组</li> <li>公网NAT网关</li> <li>虚拟私有云</li> <li>负载均衡（ELB）</li> </ul> |
| 七牛云 | 存储   | 对象存储                                                                                                                                  |
| 金山云 | 存储   | 对象存储                                                                                                                                  |

**步骤5** 单击“确认”，在线调研任务创建完成，进入任务中心页面，系统自动开始资源发现。

- 当任务状态为“成功”时，返回资源采集页面，[查看发现的源端资源列表和资源详情](#)。
- 当任务状态为“失败”时，单击任务名称，查看采集失败的数据源，将鼠标放到失败数据源的采集状态上，可以查看失败原因。

图 1-19 查看失败原因



**步骤6** 任务状态为成功后，根据不同资源类型，进行后续操作。

- 主机资源需要先完成[深度采集](#)或[源端迁移准备度检查](#)，再进行[迁移方案设计](#)或[迁移方案配置](#)。
- 容器资源需要先完成[深度采集](#)，再进行[迁移方案设计](#)。
- 中间件、数据库、存储等资源可直接进行[迁移方案设计](#)。支持对AWS RDS（包括MySQL、MariaDB、Aurora、PGSQL、SQLServer、Oracle）和AWS DocumentDB数据库进行[深度采集](#)，以获取数据库更详细的信息。
- 对象存储资源可以进行[深度采集](#)和[存储迁移方案配置](#)。

----结束

## 主机源端迁移准备度检查

请按照以下步骤对主机资源进行迁移预检查（源端迁移准备度），确定源端主机是否满足迁移条件。

- 在源端内网环境中[安装MgC Agent](#)并与迁移中心连接。
- 在资源采集页面的主机资源列表，勾选待迁移的主机，单击列表上方的[关联到应用](#)按钮，弹出关联到应用窗口。如果在创建在线调研任务时，已经配置应用参数，可以跳过该步骤。

- 已提前创建应用，在下拉列表中选择要关联的应用，单击“确定”。
  - 未提前创建应用，单击“新建应用”，输入自定义的应用名称和描述，选择业务场景、使用场景和目的端区域，单击“创建”按钮，应用创建成功。选择创建的应用后，单击“确定”。
3. 在列表上方选择“目标场景>主机迁移”，此时源端迁移准备度列会出现“配置”按钮。
  4. 单击源端迁移准备度列的“配置”，弹出配置窗口。
  5. 根据表1-10，配置参数。

表 1-10 迁移准备度参数配置说明

| 参数          | 配置说明                                                                                                                                                                                                                                                                            |
|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 类型          | 根据实际的主机操作系统类型选择。                                                                                                                                                                                                                                                                |
| 迁移中心Agent绑定 | 选择在源端环境安装的MgC Agent。                                                                                                                                                                                                                                                            |
| 接入IP        | 选择主机接入IP，可以为公网IP，也可以是私有IP；迁移预检查通过后，该IP会作为后续主机迁移工作流的资源接入IP。                                                                                                                                                                                                                      |
| 端口          | 主机开放端口。 <ul style="list-style-type: none"> <li>• Windows主机默认为5985端口，无法修改。</li> <li>• Linux主机默认22端口，可根据实际情况修改。</li> </ul>                                                                                                                                                        |
| 凭证          | 选择主机凭证，如未提前在MgC Agent上添加源端主机凭证，请前往MgC Agent控制台添加，并同步至迁移中心。<br><b>须知</b><br>为确保采集过程具备充分的系统访问权限，从而能够获取到必要的信息和数据。对主机深度采集的凭证要求如下： <ul style="list-style-type: none"> <li>• 对Linux主机进行深度采集时，请添加Root账号和密码作为采集凭证。</li> <li>• 对Windows主机进行深度采集时，请添加Administrator账号和密码作为采集凭证。</li> </ul> |

6. 配置完成后，单击“确认”按钮，系统会校验配置信息的正确性并开始源端迁移准备度检查，当源端迁移准备度列状态会变为已就绪时，代表该主机已满足迁移条件，可以进行下一阶段迁移方案设计或配置主机迁移方案。

## 主机深度采集

请按照以下步骤对主机资源进行深度采集。

1. 在源端内网环境中安装MgC Agent并与迁移中心连接。
2. 在资源采集页面的主机资源列表，单击需要进行深度采集的主机迁移中心Agent绑定列的“配置”按钮，弹出配置窗口。根据表1-11，配置参数。

表 1-11 深度采集参数配置说明

| 参数          | 配置说明                                                                                                                                                                                                                                                                                          |
|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 类型          | 根据实际的主机操作系统类型选择。                                                                                                                                                                                                                                                                              |
| 迁移中心Agent绑定 | 选择在源端环境安装的MgC Agent。                                                                                                                                                                                                                                                                          |
| 接入IP        | 选择主机接入IP，可以为公网IP，也可以是私有IP；迁移预检查通过后，该IP会作为后续主机迁移工作流的资源接入IP。                                                                                                                                                                                                                                    |
| 端口          | 主机开放端口。 <ul style="list-style-type: none"> <li>Windows主机默认为5985端口，无法修改。</li> <li>Linux主机默认22端口，可根据实际情况修改。</li> </ul>                                                                                                                                                                          |
| 凭证          | 选择主机凭证，如未提前在MgC Agent上 <a href="#">添加源端主机凭证</a> ，请前往MgC Agent控制台添加，并同步至迁移中心。<br><b>须知</b><br>为确保采集过程具备充分的系统访问权限，从而能够获取到必要的信息和数据。对主机深度采集的凭证要求如下： <ul style="list-style-type: none"> <li>对Linux主机进行深度采集时，请添加Root账号和密码作为采集凭证。</li> <li>对Windows主机进行深度采集时，请添加Administrator账号和密码作为采集凭证。</li> </ul> |

- 单击“**确认**”，MgC Agent和凭证都绑定成功后，系统会自动开始深度采集。当深度采集列的状态为“**已采集**”时，代表采集完成。可以进行下一阶段[迁移方案设计](#)或[迁移方案配置](#)。

## 容器深度采集

按照以下步骤对**容器资源**进行深度采集。

- 在源端内网环境中**安装MgC Agent**并与迁移中心连接。
- 在**资源采集**页面，单击**容器资源**类型，进入容器资源列表。
- 在需要进行深度采集的容器**迁移中心Agent绑定**列，单击“**绑定**”，弹出**绑定采集设备**窗口。  
如果需要批量绑定，可以同时勾选多个资源，单击列表上方的“**绑定迁移中心Agent**”，弹出**绑定迁移中心Agent**窗口。
- 选择源端安装的MgC Agent，单击“**确定**”。绑定成功后，资源的采集设备列状态会变为已绑定。
- MgC Agent绑定成功后，还需要为资源绑定凭证，单击资源凭证列的“**绑定**”，弹出**绑定凭证**窗口。
- 选择资源凭证，如未提前在MgC Agent上[添加源端资源凭证](#)，请前往MgC Agent控制台添加，并同步至迁移中心。
- 单击“**确定**”，系统会自动检查凭证绑定状态。当深度采集列状态为**就绪**时，单击深度采集列的“**采集**”进行深度采集。首次深度采集完成后，可以单击采集状态列的“**重新采集**”按钮，进行多次深度采集。采集完成后，单击资源名称可以查看采集到的容器详情。

## 对象存储深度采集

通过深度采集获取对象存储资源的详细信息，以提高迁移集群规格评估的准确性。请按照以下步骤进行**对象存储资源**深度采集。

1. 在源端内网环境中**安装MgC Agent**并与迁移中心连接。
2. 在**资源采集**页面，单击**存储资源**类型，进入存储资源列表。
3. 在对象存储资源列表中，单击**深度采集配置**列的“**配置**”，弹出**配置**窗口。
4. 选择源端安装的MgC Agent和相应的资源凭证，单击“**确认**”完成配置。如未提前提前在MgC Agent上**添加源端资源凭证**，请前往MgC Agent控制台添加，并同步至迁移中心。

### 须知

在对**Azure**云平台对象存储资源进行深度采集时，需要使用“**存储账户**”和“**密钥**”作为采集凭证，凭证的获取方法请参见[如何获取Azure对象存储深度采集所需凭证](#)。

5. 配置完成后，单击操作列的“**添加前缀**”，弹出**添加前缀**窗口。
6. 输入指定的资源目录路径进行采集，不输入则默认采集全桶资源，单击“**确定**”保存前缀设置。
7. 完成深度采集配置和添加前缀后，单击操作列的“**深度采集**”，系统开始进行对象存储资源的深度采集。支持进行多次深度采集。当采集状态为**采集完成后**，单击资源名称，可以查看采集到的信息。

## 数据库深度采集

当前支持对AWS RDS（包括MySQL、MariaDB、Aurora、Postgre SQL、SQL Server、Oracle）和AWS DocumentDB数据库进行深度采集，以获取包括数据库版本、引擎、服务器字符集、平均事务每秒(Transaction Per Second, TPS)和查询每秒(Query Per Second, QPS)等关键性能指标在内的详细信息。不同的数据库类型，采集的信息详情可能有所差异。

1. 在源端内网环境中或在可以访问到源端数据库的网络环境中**安装MgC Agent**并与迁移中心连接。
2. 在**资源采集**页面，单击**数据库**资源类型，进入数据库资源列表。
3. 在数据库资源列表中，通过**平台类型**过滤出采集到的所有AWS数据库资源，在支持深度采集的数据库类型的**迁移中心Agent绑定**列，单击“**绑定**”，弹出**绑定迁移中心Agent**窗口。

如果需要批量绑定，可以同时勾选多个资源，单击列表上方的“**绑定迁移中心Agent**”，弹出**绑定迁移中心Agent**窗口。

4. 选择源端安装的MgC Agent，以及网络接入方式（如果采集的资源与安装MgC Agent的主机处于同一VPC内，可以选择私网接入，否则需要选择公网接入），单击“**确定**”。绑定成功后，资源的采集设备列状态会变为已绑定。
5. 采集设备绑定成功后，还需要为资源绑定凭证，单击**资源凭证**列的“**绑定**”，弹出**绑定凭证**窗口。
6. 选择数据库资源凭证，如未提前提前在MgC Agent上**添加源端资源凭证**，请前往MgC Agent控制台添加，并同步至迁移中心。

7. 单击“确定”，系统会自动检查凭证绑定状态。当深度采集列的状态为就绪时，单击“采集”开始深度采集。首次深度采集完成/采集失败后，可以单击深度采集列的“重新采集”，进行多次深度采集。
8. 深度采集完成后，单击数据库名称，进入数据库详情页面，在数据库详情区域，可以查看采集到的详细信息。

## 查看发现的源端资源列表和资源详情

发现任务仅能获取源端资源的基本信息，更多详细信息需要通过深度采集获取。

1. 在资源采集页面，可以查看发现的各种类型资源列表和资源数量。
2. 单击资源类型，进入对应的资源列表。
3. 单击资源名称，可以查看该资源的基本信息。

## 1.4.2 内网采集

适用于采集线下自建IDC机房主机和vCenter上的主机信息。通过在源端内网环境中部署MgC Agent，使用网段扫描或VMware扫描的方式，发现和采集源端主机资源。

### 使用须知

- 目前支持采集的VMware版本包括：**VMware 5.0~7.0**
- VMware扫描和网段扫描两种方法都采用私有IP结合MgC Agent ID的方式来标识内网中的主机资源。这意味着，一旦完成一次资源采集，如果被采集的主机私有IP发生变更，系统将在下一次采集时将其识别为一个新的主机资源，从而导致主机资源数量增加。为了避免这种情况，建议在迁移过程完成之前，不要更改源端主机的私有IP地址。
- 通过VMware扫描发现的主机支持自动采集指定周期内的主机性能数据，无需再手动配置进行性能采集。
- 通过VMware扫描方式采集的性能数据来源于vCenter中留存的历史样本统计数据。vCenter历史样本统计数据的默认留存策略为：
  - 最近1天：每5分钟留存一条数据。
  - 最近1周：每30分钟留存一条数据。
  - 最近1月：每2小时留存一条数据。

### 前提条件

- 已在源端内网环境**安装MgC Agent**，并与**迁移中心连接**。
- 已在MgC Agent上**添加源端主机凭证**。

#### 须知

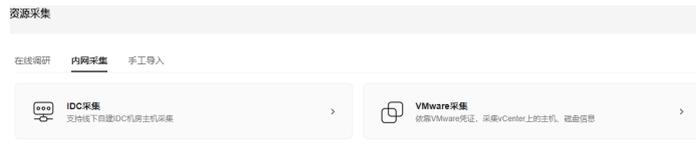
为确保采集过程具备充分的系统访问权限，从而能够获取到必要的信息和数据。对主机深度采集的凭证要求如下：

- 对Linux主机进行深度采集时，请添加Root账号和密码作为采集凭证。
- 对Windows主机进行深度采集时，请添加Administrator账号和密码作为采集凭证。

## 创建 IDC 采集任务

- 步骤1** 登录**迁移中心管理控制台**。在左侧导航栏的项目下拉列表中选择创建的**应用迁移项目**。
- 步骤2** 单击左侧导航栏的“**源端发现>资源采集**”，进入**资源采集**页面。
- 步骤3** 在“**内网采集**”页签，根据实际情况，选择源端主机资源所部署的环境。

图 1-20 选择主机部署的环境



- 步骤4** 选择IDC采集时，参考**表1-12**，进行配置。

表 1-12 IDC 采集参数配置说明

| 参数        | 说明                                                                                                                                                                                                                      |
|-----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 任务名称      | 自定义任务名称。                                                                                                                                                                                                                |
| 任务说明      | 输入任务说明。                                                                                                                                                                                                                 |
| 迁移中心Agent | 选择在源端内网环境中安装的 <b>MgC Agent</b> 。                                                                                                                                                                                        |
| 协议        | 只支持TCP协议。                                                                                                                                                                                                               |
| 网段        | 输入IP地址段，需在以下范围内： <ul style="list-style-type: none"> <li>10.0.0.0 ~10.255.255.255</li> <li>172.16.0.0 ~172.31.255.255</li> <li>192.168.0.0 ~ 192.168.255.255</li> </ul>                                                  |
| Linux     | 填写Linux系统类型主机需要扫描的端口。如果不需要扫描Linux类型主机，可将端口设为0。                                                                                                                                                                          |
| Windows   | 填写Windows系统类型主机需要扫描的端口。如果不需要扫描Windows类型主机，可将端口设为0。                                                                                                                                                                      |
| 应用        | 可选参数，选择采集到的资源要加入的应用，通过应用可以将资源进行分组管理。如果下拉列表中无应用，按如下步骤创建： <ol style="list-style-type: none"> <li>单击“<b>新建应用</b>”，输入自定义的<b>应用名称和描述</b>；选择<b>业务场景和使用场景</b>；<b>目的端区域</b>选择目的端所在区域。</li> <li>单击“<b>确定</b>”按钮，完成创建。</li> </ol> |

- 步骤5** 单击“**确认**”，IDC采集任务创建完成，进入任务中心页面，系统自动开始主机资源发现。
- 步骤6** 采集任务状态为成功后，返回**资源采集**页面，进行主机深度采集。
- 在**资源采集**页面的主机资源列表。单击需要进行深度采集的主机**迁移中心Agent**绑定列的“**配置**”按钮，弹出**配置**窗口。

2. 根据表1-13，配置参数。

表 1-13 深度采集参数配置说明

| 参数          | 配置说明                                                                |
|-------------|---------------------------------------------------------------------|
| 类型          | 根据实际的主机操作系统类型选择。                                                    |
| 迁移中心Agent绑定 | 选择在源端环境安装的MgC Agent。                                                |
| 接入IP        | 选择主机接入IP，可以为公网IP，也可以是私有IP；迁移预检查通过后，该IP会作为后续主机迁移工作流的资源接入IP。          |
| 端口          | 主机开放端口。<br>- Windows主机默认为5985端口，无法修改。<br>- Linux主机默认22端口，可根据实际情况修改。 |
| 凭证          | 选择主机凭证，如未提前在MgC Agent上添加源端主机凭证，请前往MgC Agent控制台添加，并同步至迁移中心。          |

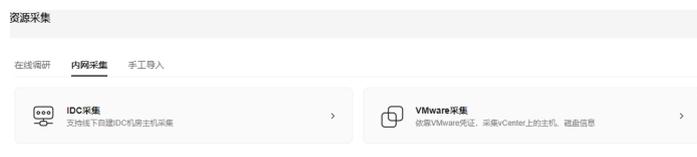
3. 单击“确认”，MgC Agent和凭证都绑定成功后，系统会自动开始深度采集。当深度采集列的状态为“已采集”时，代表采集完成。可以进行下一阶段迁移方案设计或配置主机迁移方案。

---结束

## 创建 VMware 采集任务

- 步骤1** 登录迁移中心管理控制台。在左侧导航栏的项目下拉列表中选择创建的应用迁移项目。
- 步骤2** 单击左侧导航栏的“源端发现>资源采集”，进入资源采集页面。
- 步骤3** 在“内网采集”页签，根据实际情况，选择源端主机资源所部署的环境。

图 1-21 选择主机部署的环境



- 步骤4** 选择VMware采集时，参考表1-14，进行配置。

表 1-14 VMware 采集参数配置说明

| 参数   | 说明       |
|------|----------|
| 任务名称 | 自定义任务名称。 |
| 任务说明 | 输入任务说明。  |

| 参数        | 说明                                                                                                                                                                                                                                                 |
|-----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 迁移中心Agent | 选择在源端内网环境中安装的MgC Agent。                                                                                                                                                                                                                            |
| 连接地址      | 填写vCenter的IP地址。                                                                                                                                                                                                                                    |
| 端口        | 填写访问vCenter的端口。                                                                                                                                                                                                                                    |
| 凭证        | 选择vCenter的凭证。如下拉列表为空，请前往MgC Agent <a href="#">添加私有云凭证</a> 。凭证中的用户名、密码为vCenter的登录用户名和密码。                                                                                                                                                            |
| 应用        | <p>可选参数，选择采集到的资源要加入的应用，通过应用可以将资源进行分组管理。如果下拉列表中无应用，按如下步骤创建：</p> <ol style="list-style-type: none"> <li>1. 单击“<b>新建应用</b>”，输入自定义的<b>应用名称</b>和<b>描述</b>；选择<b>业务场景</b>和<b>使用场景</b>；<b>目的端区域</b>选择目的端所在区域。</li> <li>2. 单击“<b>确定</b>”按钮，完成创建。</li> </ol> |

**步骤5** 单击“**确认**”，VMware采集任务创建完成，进入任务中心页面，系统自动开始主机资源发现。

**步骤6** 采集任务状态为成功后，返回**资源采集**页面，进行主机深度采集。

1. 在**资源采集**页面的主机资源列表。单击需要进行深度采集的主机**迁移中心Agent**绑定列的“**配置**”按钮，弹出**配置**窗口。
2. 根据**表1-15**，配置参数。

**表 1-15** 深度采集参数配置说明

| 参数          | 配置说明                                                                                                                            |
|-------------|---------------------------------------------------------------------------------------------------------------------------------|
| 类型          | 根据实际的主机操作系统类型选择。                                                                                                                |
| 迁移中心Agent绑定 | 选择在源端环境安装的MgC Agent。                                                                                                            |
| 接入IP        | 选择主机接入IP，可以为公网IP，也可以是私有IP；迁移预检查通过后，该IP会作为后续主机迁移工作流的资源接入IP。                                                                      |
| 端口          | <p>主机开放端口。</p> <ul style="list-style-type: none"> <li>- Windows主机默认为5985端口，无法修改。</li> <li>- Linux主机默认22端口，可根据实际情况修改。</li> </ul> |
| 凭证          | 选择主机凭证，如未提前在MgC Agent上 <a href="#">添加源端主机凭证</a> ，请前往MgC Agent控制台添加，并同步至迁移中心。                                                    |

3. 单击“**确认**”，MgC Agent和凭证都绑定成功后，系统会自动开始深度采集。当深度采集列的状态为“**已采集**”时，代表采集完成。可以进行下一阶段**迁移方案设计**或**配置主机迁移方案**。

---**结束**

## 1.4.3 导入 RVTools 资源

### 约束与限制

- **版本要求**  
支持导入RVTools以下版本的导出结果：
  - 4.4.1
  - 4.4.2
  - 4.4.3
  - 4.4.4
  - 4.4.5
  - 4.5.0
  - 4.5.1
  - 4.6.1
- **文件格式要求**  
导出结果的文件格式必须为Excel (.xlsx)。
- **文件大小、压缩比要求**  
导入的文件大小不能超过100 MB，压缩比不能低于5%。

### 步骤一：导出 RVTools 资源

**步骤1** 启动RVTools应用程序，进入登录界面。

**步骤2** 在登录界面输入以下信息。

- 在IP地址栏中，输入vCenter服务器的IP地址。
- 在用户名栏中，填写用于连接vCenter的用户名。
- 在密码栏中，输入与用户名相对应的密码。

**步骤3** 单击“**登录**”按钮。登录成功后，在菜单栏单击“**file > Export all to Excel**”。

**步骤4** 选择保存路径，将资源导出到本地。

----结束

### 步骤二：将导出的文件上传至 OBS

**步骤1** 登录华为云**OBS控制台**，创建一个用于存储RVTools导出资源文件的**标准存储桶**。方法请参考**创建桶**。

**步骤2** 参考**上传对象**，将**导出的RVTools资源**文件上传至OBS标准存储桶。

----结束

### 步骤三：导入 RVTools 资源

**步骤1** 登录**迁移中心管理控制台**。在左侧导航栏的项目下拉列表中选择创建的**应用迁移项目**。

**步骤2** 单击左侧导航栏的“源端发现>资源采集”，进入资源采集页面。

**步骤3** 在手工导入页签，单击“手工导入”，弹出手工导入窗口。

图 1-22 手工导入



**步骤4** 选择“RVTools导入”，在文件的下拉列表中，找到**步骤二**存放RVTools导出资源文件的OBS桶，并选择导出的资源文件，单击“确认”按钮，上传至迁移中心。

图 1-23 上传 RVTools 导出文件



### 说明

首次使用该功能时需要授权委托，单击RVTools导入委托窗口的“同意授权”即可。

**步骤5** 上传成功后，进入任务中心页面，查看任务状态。

- 当任务状态为“成功”时，您可以查看导入的[资源详情](#)和[资源统计信息](#)。
- 如果任务状态为“失败”，单击任务名称，可以查看采集失败的数据源，将鼠标放到数据源的采集状态上，可以查看失败原因。常见失败原因的处理方法请参考[导入RVTools资源任务失败](#)。处理完失败原因后，重新导入文件进行采集。

----结束

## 查看主机资源详情

当RVTools导入任务状态为“成功”时，在资源采集页面的主机资源列表，可以查看导入的主机资源信息，单击主机资源名称，可以查看该主机的更多基本信息、规格信息等。

## 查看资源统计信息

当RVTools导入任务状态为“成功”时，在资源统计页面，可以查看VMware资源的统计信息，包括：

- 宿主机：CPU类型统计、宿主机个数统计、内存总量统计、已分配虚拟机内存、已分配占比、虚拟机总数统计、虚拟机运行总数统计、存储使用详情统计以及主机列表详情。

- 虚拟机：Windows操作系统个数统计、Linux系统个数统计、内存使用总容量、CPU使用核数、磁盘使用个数、磁盘容量占比、虚拟机数量、虚拟机列表、操作系统。

**步骤1** 在资源采集页面，单击页面右上角的“VMware资源”，进入资源统计页面。

图 1-24 查看 VMware 资源



**步骤2** 分别选择VMware资源的“宿主机”页签和“虚拟机”页签，查看对应的统计信息。

---结束

## 1.4.4 导入阿里云资源清单

从阿里云导出资源清单（CSV格式），再将其导入MgC，系统会自动解析这些文件，识别出相应的资源，并在资源列表展示资源的详细信息。此外，利用解析后的信息，您还可以使用迁移中心的规格评估功能，为您推荐与阿里云资源相匹配的华为云资源规格。

支持导入的阿里云资源类型包括：

- 云服务器 ECS实例和云盘
- 对象存储 OSS桶
- 云数据库 RDS MySQL实例
- 云数据库 Redis实例

### 导出阿里云 ECS 实例和云盘

**步骤1** 登录阿里云ECS控制台。

**步骤2** 在概览页面，单击“导出数据>实例”，导出所有实例资源清单。



**步骤3** 单击“导出数据>云盘”，导出所有云盘资源清单。

**步骤4** 打开实例资源清单和云盘资源清单，根据实际需求，调整清单内的资源信息。请确保清单中包含了以下必要信息：

| 清单类型 | 必选字段                   |
|------|------------------------|
| 实例清单 | 实例ID、操作系统、实例规格、CPU、内存  |
| 云盘清单 | 实例Id、云盘Id、容量（GiB）、云盘属性 |

----结束

## 导出阿里云 OSS 桶

**步骤1** 登录阿里云OSS控制台。

**步骤2** 在左侧导航栏选择“**Bucket 列表**”，进入**Bucket 列表**页面。

**步骤3** 单击列表右上角的  图标，弹出**导出Bucket列表**窗口。



**步骤4** 勾选需要导出的信息，单击“**确定**”，导出所有Bucket资源清单。

----结束

## 导出阿里云 RDS MySQL 实例

**步骤1** 登录阿里云RDS控制台。

**步骤2** 在左侧导航栏选择“**实例列表**”，进入**实例列表**页面。在页面上方选择地域。

**步骤3** 通过**数据库类型**列过滤出所有MySQL类型的实例，单击列表右上角的  图标。



**步骤4** 在弹出的对话框中，勾选需要导出的信息。单击“**确定**”。系统会将实例信息输出至CSV格式的文件并自动下载。

**步骤5** 由于从阿里云导出的CSV文件内容格式不满足MgC解析标准，因此需要进行如下修改：

1. 使用文件编辑器打开下载的资源清单CSV文件，在表头行搜索“PrivatePort(A network port,)”。
2. 将该表头名中的英文逗号删除，然后保存文件。正确的表头名应该为“PrivatePort(A network port)”。

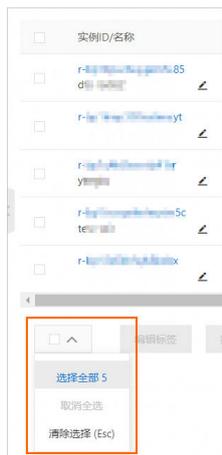
----结束

## 导出阿里云 Redis 实例

**步骤1** 登录阿里云云数据库 Redis 版控制台。

**步骤2** 在左侧导航栏选择“**实例列表**”，进入**实例列表**页面。在页面上方选择地域。

**步骤3** 在实例列表页面左下方，展开  ，单击“**选择全部**”。



**步骤4** 单击列表右上角的  图标，在弹出的对话框中，勾选需要导出的信息。



**步骤5** 单击“确定”。系统会将实例信息输出至CSV格式的文件并自动下载。

**步骤6** 由于从阿里云导出的CSV文件内容格式不满足MgC解析标准，因此需要进行如下修改：

1. 使用文件编辑器打开下载的资源清单CSV文件，
2. 在表头行的末尾，添加一个英文逗号，然后保存文件。

例如，原始表头为：

ID,Name,Quantity

修改后的表头应该是：

ID,Name,Quantity,

----结束

## 导入阿里云资源清单

**步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择创建的[应用迁移项目](#)。

**步骤2** 单击左侧导航栏的“源端发现>资源采集”，进入资源采集页面。

**步骤3** 在手工导入页签，单击“手工导入”，弹出手工导入窗口。

图 1-25 手工导入



**步骤4** 选择“**阿里云资源清单导入**”，在资源类型下拉列表中，选择要导入的资源类型，并上传对应的资源清单。

**图 1-26** 选择资源类型



### 须知

仅支持上传UTF-8编码且为中文表头的CSV文件，单个文件大小不能超过50 MB。将CSV文件编码转换为UTF-8的方法请参考[转换CSV文件编码为UTF-8](#)。

| 资源类型  | 操作步骤                                                                                                                                                                                   |
|-------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 主机    | <ol style="list-style-type: none"> <li>单击<b>资源清单</b>后的“<b>添加文件</b>”，将<b>导出的阿里云ECS实例资源清单</b>上传。</li> <li>实例资源清单上传后，单击<b>云盘资源清单</b>后的“<b>添加文件</b>”，将<b>导出的阿里云ECS云盘资源清单</b>上传。</li> </ol> |
| Redis | 单击 <b>资源清单</b> 后的“ <b>添加文件</b> ”，将 <b>导出的阿里云Redis实例资源清单</b> 上传。                                                                                                                        |
| 对象存储  | 单击 <b>资源清单</b> 后的“ <b>添加文件</b> ”，将 <b>导出的阿里云OSS桶资源清单</b> 上传。                                                                                                                           |
| 数据库   | 单击 <b>资源清单</b> 后的“ <b>添加文件</b> ”，将 <b>导出的阿里云RDS MySQL实例资源清单</b> 上传。                                                                                                                    |

**步骤5** 单击“**确认**”，进入任务中心页面，资源清单上传成功后，返回**资源采集**页面，在对应的资源类型列表查看导入的资源。

**步骤6** （可选）将导入的阿里云资源**关联应用**，并使用**规格评估**功能，获取与阿里云资源相匹配的华为云资源规格。

----**结束**

## 1.4.5 导入应用关联文件

将您源端配置管理数据库中的应用关联信息，填入导入模板表格中，再将表格导入迁移中心，分析出应用间的依赖及关联关系。

### 操作步骤

**步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择创建的[应用迁移项目](#)。

**步骤2** 单击左侧导航栏的“[源端发现>资源采集](#)”，进入[资源采集](#)页面。

**步骤3** 在[手工导入](#)页签，单击“[手工导入](#)”，弹出手工导入窗口。

图 1-27 手工导入



**步骤4** 选择“[应用关联文件导入](#)”，单击“[下载模板](#)”，将导入模板下载到本地。

图 1-28 下载应用关联模板



**步骤5** 打开导入模板，填写模板内的参数，其中标黄的参数为必填。模板参数填写完成并保存后，单击导入窗口的“[添加文件](#)”，将填写后的模板文件上传至迁移中心。

**步骤6** 上传成功后，单击“[确认](#)”按钮，进入任务中心页面，系统自动开始采集。

- 当任务状态为“**成功**”时，您可以前往[应用管理](#)页面，查看导入结果。
- 如果任务状态为“**失败**”，单击任务名称，可以查看采集失败的数据源，将鼠标放到数据源的采集状态上，可以查看失败原因。处理完失败原因后，重新导入文件进行采集。

----结束

## 1.4.6 导入工具采集结果

通过线下采集器，采集本地资源信息，并将采集结果上传到迁移中心进行应用关联分析。

## 前提条件

已[创建工具采集任务](#)，并获取采集结果。

## 操作步骤

- 步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择创建的[应用迁移项目](#)。
- 步骤2** 单击左侧导航栏的“源端发现>资源采集”，进入资源采集页面。
- 步骤3** 在手工导入页签，单击“手工导入”，弹出手工导入窗口。

图 1-29 手工导入



- 步骤4** 选择“工具采集结果导入”，单击导入窗口的“添加文件”，将工具采集结果的json文件上传至迁移中心。

图 1-30 添加工具采集结果文件



- 步骤5** 上传成功后，单击“确认”按钮，进入任务中心页面，系统自动开始采集。
  - 当任务状态为“成功”时，您可以前往[应用管理](#)页面，查看导入结果。
  - 如果任务状态为“失败”，单击任务名称，可以查看采集失败的数据源，将鼠标放到数据源的采集状态上，可以查看失败原因。处理完失败原因后，重新导入文件进行采集。

----结束

## 1.4.7 手动添加资源

目前支持手动添加的资源类型包括：**主机**、**Oracle数据库**。通过在源端环境中安装MgC Agent（原Edge）进行资源采集。

## 准备工作

- 在MgC控制台[创建项目](#)。

- 在源端内网环境[安装MgC Agent](#)，并与[迁移中心连接](#)。
- 在MgC Agent上[添加源端资源凭证](#)。

## 添加主机

- 步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择创建的[应用迁移项目](#)。
- 步骤2** 单击左侧导航栏的“[源端发现>资源采集](#)”，进入[资源采集](#)页面。
- 步骤3** 在主机资源列表，单击列表上方的“[添加资源](#)”按钮，弹出添加主机窗口。

图 1-31 添加主机



- 步骤4** 根据[表1-16](#)，填写相应参数后，单击“[确认](#)”，系统会自动检查凭证绑定状态和进行信息采集。

表 1-16 添加主机参数说明

| 参数          | 说明                                                                                                                       |
|-------------|--------------------------------------------------------------------------------------------------------------------------|
| 名称          | 用户自定义                                                                                                                    |
| 迁移中心Agent绑定 | 选择源端注册的MgC Agent。                                                                                                        |
| 类型          | 主机实际系统类型                                                                                                                 |
| 接入IP        | 填写主机IP。<br>如果采集的主机与安装MgC Agent的主机处于同一VPC内，可以填写私有IP，否则需要填写公网IP。                                                           |
| 端口          | 主机开放端口。 <ul style="list-style-type: none"> <li>• Windows主机默认为5985端口，无法修改。</li> <li>• Linux主机默认22端口，可根据实际情况修改。</li> </ul> |
| 凭证          | 选择源端主机凭证，如未提前在MgC Agent上 <a href="#">添加源端主机凭证</a> ，请前往MgC Agent控制台添加，并同步至迁移中心。                                           |

**步骤5** 添加成功后，在主机资源列表，可以查看添加的主机资源。

----结束

## 添加数据库

**步骤1** 登录**迁移中心管理控制台**。在左侧导航栏的项目下拉列表中选择创建的**应用迁移项目**。

**步骤2** 单击左侧导航栏的“**源端发现>资源采集**”，进入**资源采集**页面。

**步骤3** 单击数据库资源类型，进入数据库资源列表。

**步骤4** 单击“**添加资源**”按钮，弹出**添加数据库**窗口。

**图 1-32** 添加数据库资源



**步骤5** 根据**表1-17**，填写相应参数。

添加数据库
×

\* 名称

\* 迁移中心Agent绑定

\* 类型

\* IP   
请确保迁移中心Agent能通过该地址连接到该资源。

\* 端口

\* 服务名称/SID

\* 凭证

**表 1-17** 添加数据库参数配置说明

| 参数 | 说明    |
|----|-------|
| 名称 | 用户自定义 |

| 参数          | 说明                                                                                                                   |
|-------------|----------------------------------------------------------------------------------------------------------------------|
| 迁移中心Agent绑定 | 选择在源端注册并与迁移中心连接的MgC Agent。                                                                                           |
| 类型          | 选择“Oracle”。                                                                                                          |
| IP          | 输入可以连接到源端Oracle数据库服务器的IP地址。                                                                                          |
| 端口          | 源端Oracle 数据库监听端口。默认为1521。                                                                                            |
| 服务名称/SID    | 源端数据库待采集的服务名。默认为ORCL。<br>您可以登录Oracle 数据库，使用如下命令，查看并确定需要采集的服务名称。<br><pre>SELECT instance_name FROM v\$instance;</pre> |
| 凭证          | 选择源端数据库凭证，如未提前在MgC Agent上 <a href="#">添加源端数据库凭证</a> ，请前往MgC Agent控制台添加，并同步至迁移中心。资源类型选择“数据库”，认证方式选择“用户名/密码”。          |

**步骤6** 参数配置完成后，单击“确认”，系统会自动进行深度采集。当深度采集列的状态变为**已采集**，代表采集完成。单击数据库名称，进入数据库详情页面，可以查看采集到的信息。

---结束

## 1.4.8 采集主机性能

迁移中心支持对主机资源进行性能数据采集，获取源端主机的CPU、内存负载、磁盘IOPS，内网出入流量，网络出入包速率，网络连接数等性能信息。在后续进行目的端规格评估时，可以根据采集到的源端主机实际性能负载推荐目的端主机CPU核数、内存大小等性能规格。

### 使用须知

- 通过VMware扫描发现的主机会自动采集指定周期内的主机性能数据，无需再配置性能采集。VMware扫描方式采集的性能数据来源于vCenter中留存的历史样本统计数据。vCenter历史样本统计数据的默认留存策略为：
  - 最近1天：每5分钟留存一条数据。
  - 最近1周：每30分钟留存一条数据。
  - 最近1月：每2小时留存一条数据。
- 开始采集后，默认每5分钟进行一次数据采集。
- 建议采集时长≥1小时，若采集时长不足，将无法获取7天和30天最大值以及平均值数据。
- 采集周期默认为7天，7天后会自动停止采集；采集周期内可以进行停止采集和重启采集操作；重启采集后，采集周期重新计算。
- 多种方式采集的性能数据重叠时，以最新数据为准。
- 采集到的性能数据将保留180天，超期后会自动删除。
- MgC Agent（原Edge）建议安装在独立主机上，避免与业务主机共用，以确保采集的稳定性和效率。

- Linux主机：MgC Agent最多同时采集1000台Linux主机。当同时采集1000台Linux主机时，安装MgC Agent的主机规格最低要求为8U16G，且可使用内存不小于8G。
- Windows主机：MgC Agent最多同时采集500台Windows主机，当同时采集500台Windows主机时，安装MgC Agent的主机规格最低要求为16U32G，且可使用内存不小于8G。

## 前提条件

- 已通过[在线调研/内网采集/手动添加](#)等方式完成主机资源采集。
- 已在源端内网环境[安装MgC Agent](#)，并与[迁移中心连接](#)。
- 已在MgC Agent上[添加源端主机凭证](#)。

## 操作步骤

**步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择创建的[应用迁移项目](#)。

**步骤2** 单击左侧导航栏的“[源端发现>资源采集](#)”，进入[资源采集](#)页面。

**步骤3** 在主机资源列表，为需要进行性能采集的主机绑定MgC Agent和凭证。

- 如果性能采集列的“开始”按钮置灰，将鼠标移动到“开始”按钮处，在弹出的提示框中，单击“去配置”，弹出配置窗口。



- 如果性能采集列的“开始”按钮未置灰，代表已完成配置，可以跳过此步执行[步骤7](#)。

**步骤4** 根据[表1-18](#)，配置参数。

表 1-18 迁移准备度参数配置说明

| 参数          | 配置说明                                                                                                                     |
|-------------|--------------------------------------------------------------------------------------------------------------------------|
| 类型          | 根据实际的主机操作系统类型选择。                                                                                                         |
| 迁移中心Agent绑定 | 选择在源端环境安装的MgC Agent。                                                                                                     |
| 接入IP        | 选择主机接入IP，可以为公网IP，也可以是私有IP；在需要使用代理连接主机的情况下，也可以自行输入接入IP。迁移预检查通过后，该IP会作为后续主机迁移工作流的资源接入IP。                                   |
| 端口          | 主机开放端口。 <ul style="list-style-type: none"> <li>• Windows主机默认为5985端口，无法修改。</li> <li>• Linux主机默认22端口，可根据实际情况修改。</li> </ul> |
| 凭证          | 选择主机凭证，如未提前在MgC Agent上 <a href="#">添加源端主机凭证</a> ，请前往MgC Agent控制台添加，并同步至迁移中心。                                             |

**步骤5** 配置完成后，单击“[确认](#)”按钮，系统会校验配置信息的正确性。

**步骤6** 配置信息校验无误后，单击性能采集列的“开始”，开始主机性能采集，此时状态为“采集中”。

开始性能采集后，您还可以进行如下操作：

- **查看采集数据**

- a. 单击主机名称，进入主机详情页面，在性能信息区域，可以查看采集到的主机性能数据。

---

**须知**

- 通过性能采集任务采集性能数据时，性能采集停止后才会开始计算7天和30天的聚合值。
- 如果采集时间小于1小时，则采集停止后1小时才会开始计算。
- 多种方式采集的性能数据重叠时，以最新数据为准。

- b. 单击操作列的“查看”，可以查看采集周期内每项数据的详细数值和曲线图。

- **停止采集**

单击“停止”，可以暂停性能数据采集。

---结束

## 1.4.9 关联应用

### 概述

通过应用将资源进行分组管理，便于后续进行规格评估和创建迁移 workflow。

### 操作步骤

**步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择创建的[应用迁移项目](#)。

**步骤2** 单击左侧导航栏的“源端发现>资源采集”，进入资源采集页面。

**步骤3** 在资源列表选择资源类型，进入对应的资源列表页面。

**步骤4** 勾选需要关联到同一应用的资源，单击页面左上角的“关联到应用”，弹出关联到应用窗口。

**步骤5** 在下拉列表中选择需要关联的应用。未提前创建应用，单击“新建应用”，输入自定义的应用名称和描述；根据实际需求，选择业务场景、使用场景以及目的端区域，单击“创建”按钮，应用创建成功。

**步骤6** 选择应用后，单击“确定”。在各资源的已关联应用列，查看资源所关联应用。

---结束

## 1.5 应用管理

### 1.5.1 创建应用

您可以通过创建应用，将采集到的资源添加到应用中，对资源进行功能分组，便于后续进行规格评估和创建迁移 workflow。

#### 新建应用

**步骤1** 登录**迁移中心管理控制台**。在左侧导航栏的项目下拉列表中选择**迁移项目**。

**步骤2** 单击左侧导航栏的“源端发现>应用管理”，进入**应用管理**页面。

**步骤3** 单击应用列表中的“**新建应用**”按钮，右侧弹出**新建应用**窗口。

图 1-33 新建应用



**步骤4** 输入自定义的**应用名称**和**描述**；选择**业务场景**和**使用场景**；目的端区域选择目的端所在区域，单击“**创建应用**”按钮，应用创建成功，进入**关联资源**页面。

- 如果已经进行过资源采集，并且需要将采集到的资源关联到创建的应用，请勾选需要关联到应用的资源，单击“**关联到应用**”按钮，完成应用创建和资源关联。
- 如果还未进行资源采集或暂不需要关联资源，单击“**暂不关联资源**”按钮，完成应用创建。后续在**资源列表**中也可以进行关联操作。

----结束

#### 管理应用

您可以根据不同的业务领域将应用进行分类管理。

**步骤1** 在**应用管理**页面，单击业务领域后的“**查看**”按钮，右侧弹出**业务领域**窗口。



**步骤2** 单击“**添加**”按钮，输入业务领域名称，选择该业务包含的应用，单击“**保存**”。



----结束

## 修改应用

**步骤1** 在应用列表，单击应用操作列的“修改”，弹出修改应用窗口。

**步骤2** 根据实际情况，修改应用参数后，单击“确定”，完成应用修改。

----结束

## 1.5.2 关联资源

本节介绍为应用关联资源的方法。通过应用将资源进行分组管理，为后续规格评估、创建迁移 workflow 作准备。

### 前提条件

已通过资源采集中的**在线调研/内网发现/手工导入/手动添加**等方式完成主机资源的发现。

### 操作步骤

**步骤1** 登录**迁移中心管理控制台**。在左侧导航栏的项目下拉列表中选择**迁移项目**。

**步骤2** 单击左侧导航栏的“**源端发现>应用管理**”，进入**应用管理**页面。

**步骤3** 在应用列表，单击应用名称，进入应用详情页面。

**步骤4** 选择要关联应用的资源类型。以下步骤以关联主机资源为例。其它资源类型的关联方法与主机资源的方法相同。

**步骤5** 在主机资源列表，单击列表上方的“**关联资源**”按钮，右侧弹出**关联资源**窗口。

**步骤6** 勾选需要关联到同一应用的主机资源，单击窗口右下角的“**关联到应用**”。在资源列表中即可看到关联到应用的主机资源。

----结束

## 1.5.3 手动添加资源

支持在应用内手动添加**主机**和**Oracle数据库**两种资源类型。通过在源端环境中安装 MgC Agent（原Edge）进行资源采集。

## 准备工作

- 在MgC控制台[创建项目](#)。
- 在源端内网环境[安装MgC Agent](#)，并与[迁移中心连接](#)。
- 在MgC Agent上[添加源端资源凭证](#)。
- 已在应用管理页面[创建应用](#)。

## 添加主机

- 步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择[迁移项目](#)。
- 步骤2** 单击左侧导航栏的“[源端发现>应用管理](#)”，进入[应用管理](#)页面。
- 步骤3** 在应用列表，单击应用名称，进入应用详情页面。
- 步骤4** 在主机资源列表，单击列表上方的“[添加资源](#)”按钮，弹出添加主机窗口。
- 步骤5** 根据[表1-19](#)，填写相应参数后，单击“[确认](#)”，系统会自动检查凭证绑定状态和进行信息采集。

表 1-19 添加主机参数说明

| 参数          | 说明                                                                                                                    |
|-------------|-----------------------------------------------------------------------------------------------------------------------|
| 名称          | 用户自定义                                                                                                                 |
| 迁移中心Agent绑定 | 选择源端注册的MgC Agent。                                                                                                     |
| 类型          | 主机实际系统类型                                                                                                              |
| 接入IP        | 填写主机IP。<br>如果采集的主机与安装MgC Agent的主机处于同一VPC内，可以填写私有IP，否则需要填写公网IP。                                                        |
| 端口          | 主机开放端口。 <ul style="list-style-type: none"><li>• Windows主机默认为5985端口，无法修改。</li><li>• Linux主机默认22端口，可根据实际情况修改。</li></ul> |
| 凭证          | 选择源端主机凭证，如未提前在MgC Agent上 <a href="#">添加源端主机凭证</a> ，请前往MgC Agent控制台添加，并同步至迁移中心。                                        |

- 步骤6** 添加成功后，在主机资源列表，可以查看添加的主机资源。

----结束

## 添加数据库

- 步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择[迁移项目](#)。
- 步骤2** 单击左侧导航栏的“[源端发现>应用管理](#)”，进入[应用管理](#)页面。
- 步骤3** 在应用列表，单击应用名称，进入应用详情页面。
- 步骤4** 在数据库资源列表，单击列表上方的“[添加资源](#)”按钮，弹出添加数据库窗口。

**步骤5** 根据表1-20，填写相应参数。

添加数据库
✕

\* 名称

\* 迁移中心Agent绑定

\* 类型

\* IP   
请确保迁移中心Agent能通过该地址连接到该资源。

\* 端口

\* 服务名称/SID

\* 凭证

取消
确认

**表 1-20** 添加数据库参数配置说明

| 参数          | 说明                                                                                                                   |
|-------------|----------------------------------------------------------------------------------------------------------------------|
| 名称          | 用户自定义                                                                                                                |
| 迁移中心Agent绑定 | 选择在源端注册并与迁移中心连接的MgC Agent。                                                                                           |
| 类型          | 选择“Oracle”。                                                                                                          |
| IP          | 输入可以连接到源端Oracle数据库服务器的IP地址。                                                                                          |
| 端口          | 源端Oracle 数据库监听端口。默认为1521。                                                                                            |
| 服务名称/SID    | 源端数据库待采集的服务名。默认为ORCL。<br>您可以登录Oracle 数据库，使用如下命令，查看并确定需要采集的服务名称。<br><pre>SELECT instance_name FROM v\$instance;</pre> |
| 凭证          | 选择源端数据库凭证，如未提前在MgC Agent上 <a href="#">添加源端数据库凭证</a> ，请前往MgC Agent控制台添加，并同步至迁移中心。资源类型选择“数据库”，认证方式选择“用户名/密码”。          |

**步骤6** 参数配置完成后，单击“确认”，系统会自动进行深度采集。当深度采集列的状态变为**已采集**，代表采集完成。单击数据库名称，进入数据库详情页面，可以查看采集到的信息。

----结束

## 1.5.4 采集主机性能

迁移中心支持对主机资源进行性能数据采集，获取源端主机的CPU、内存负载、磁盘IOPS，内网出入流量，网络出入包速率，网络连接数等性能信息。在后续进行目的端规格评估时，可以根据采集到的源端主机实际性能负载推荐目的端主机CPU核数、内存大小等性能规格。

本文介绍在应用详情内，对应用中的主机资源进行性能采集的操作方法。

## 使用须知

- 通过VMware扫描发现的主机会自动采集指定周期内的主机性能数据，无需再配置性能采集。VMware扫描方式采集的性能数据来源于vCenter中留存的历史样本统计数据。vCenter历史样本统计数据的默认留存策略为：  
最近1天：每5分钟留存一条数据。  
最近1周：每30分钟留存一条数据。  
最近1月：每2小时留存一条数据。
- 开始采集后，默认每5分钟进行一次数据采集。
- 建议采集时长 $\geq 1$ 小时，若采集时长不足，将无法获取7天和30天最大值以及平均值数据。
- 采集周期默认为7天，7天后会自动停止采集；采集周期内可以进行停止采集和重启采集操作；重启采集后，采集周期重新计算。
- 多种方式采集的性能数据重叠时，以最新数据为准。
- 采集到的性能数据将保留180天，超期后会自动删除。
- MgC Agent（原Edge）建议安装在独立主机上，避免与业务主机共用，以确保采集的稳定性和效率。
- Linux主机：MgC Agent最多同时采集1000台Linux主机。当同时采集1000台Linux主机时，安装MgC Agent的主机规格最低要求为8U16G，且可使用内存不小于8G。
- Windows主机：MgC Agent最多同时采集500台Windows主机，当同时采集500台Windows主机时，安装MgC Agent的主机规格最低要求为16U32G，且可使用内存不小于8G。

## 前提条件

- 已通过[关联资源](#)或[手动添加](#)等方式获取主机资源。
- 已在源端内网环境[安装MgC Agent](#)，并与迁移中心连接。
- 已在MgC Agent上[添加源端主机凭证](#)。

## 操作步骤

**步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择[迁移项目](#)。

**步骤2** 单击左侧导航栏的“[源端发现>应用管理](#)”，进入[应用管理](#)页面。

**步骤3** 在应用列表，单击应用名称，进入应用详情页面。

**步骤4** 在主机资源列表，为需要进行性能采集的主机绑定MgC Agent和凭证。

- 如果性能采集列的“开始”按钮置灰，将鼠标移动到“开始”按钮处，在弹出的提示框中，单击“去配置”，弹出配置窗口。



- 如果性能采集列的“开始”按钮未置灰，代表已完成配置，可以跳过此步执行[步骤7](#)。

**步骤5** 根据表1-21，配置参数。

**表 1-21** 迁移准备度参数配置说明

| 参数          | 配置说明                                                                                                              |
|-------------|-------------------------------------------------------------------------------------------------------------------|
| 类型          | 根据实际的主机操作系统类型选择。                                                                                                  |
| 迁移中心Agent绑定 | 选择在源端环境安装的MgC Agent。                                                                                              |
| 接入IP        | 选择主机接入IP，可以为公网IP，也可以是私有IP；在需要使用代理连接主机的情况下，也可以自行输入接入IP。迁移预检查通过后，该IP会作为后续主机迁移工作流的资源接入IP。                            |
| 端口          | 主机开放端口。 <ul style="list-style-type: none"><li>Windows主机默认为5985端口，无法修改。</li><li>Linux主机默认22端口，可根据实际情况修改。</li></ul> |
| 凭证          | 选择主机凭证，如未提前在MgC Agent上 <a href="#">添加源端主机凭证</a> ，请前往MgC Agent控制台添加，并同步至迁移中心。                                      |

**步骤6** 配置完成后，单击“**确认**”按钮，系统会校验配置信息的正确性。

**步骤7** 配置信息校验无误后，单击性能采集列的“**开始**”，开始主机性能采集，此时状态为“**采集中**”。

开始性能采集后，您还可以进行如下操作：

- **查看采集数据**

- a. 单击主机名称，进入**主机详情**页面，在**性能信息**区域，可以查看采集到的主机性能数据。

---

**须知**

- 通过性能采集任务采集性能数据时，性能采集停止后才会开始计算7天和30天的聚合值。
- 如果采集时间小于1小时，则采集停止后1小时才会开始计算。
- 多种方式采集的性能数据重叠时，以最新数据为准。

- b. 单击操作列的“**查看**”，可以查看采集周期内每项数据的详细数值和曲线图。

- **停止采集**

- a. 单击“**停止**”，可以暂停性能数据采集。

----**结束**

## 1.6 迁移准备

## 1.6.1 部署 MgC Agent (原 Edge)

使用迁移中心的大数据相关功能前，需要在可以访问源端数据来源的环境中部署 MgC Agent。

### 注意

MgC Agent不具备自动重启功能，严禁在任务执行期间重启MgC Agent，以避免任务失败。

## 准备工作

- 建议在源端内网环境中准备一台用于安装MgC Agent的Linux主机，并确保该Linux主机满足以下要求：
  - 可以[连接外网](#)，并检查是否能够访问MgC和IoTDA服务的域名，具体需要检查的域名请查看[域名](#)。
  - 检查是否已开放MgC Agent (原Edge) 需要的端口，所需端口请查看[通信矩阵](#)。
  - 安全组的出方向规则中，需要包含8883端口。
  - 操作系统为：**CentOS 8.x**
  - 推荐规格不小于**4U8G**。如果使用大数据相关功能，推荐规格不小于**8U16G**。

### 注意

不建议将MgC Agent安装在源端待迁移的主机上，主要原因包括：

- **资源消耗**：MgC Agent在采集和迁移运行过程中会消耗CPU和内存资源。在迁移任务量大时，这可能会对源端的业务运行造成影响。
  - **端口占用**：MgC Agent会占用源端主机上的一些端口，也可能对源端业务产生影响。
- 
- 检查安装MgC Agent的Linux主机是否安装了rng-tools工具，通过以下命令检查：

```
rpm -qa | grep rng-tools
```

如果以上命令未返回rng-tools的信息，代表该主机未安装rng-tools工具，请使用如下命令来安装rng-tools工具。

```
yum -y install rng-tools
```
  - 关闭安装MgC Agent的Linux主机上的杀毒、防护类软件，该类软件会拦截MgC Agent的执行，导致迁移 workflow 执行失败。
  - 在安装MgC Agent的主机安全组入网规则中，需要添加一条允许通过TCP协议访问27080端口的规则。源地址填写用于登录MgC Agent控制台的Windows主机IP地址。



- 已注册华为账号并开通华为云，并获取账号的AK/SK。
- 已在MgC控制台创建迁移项目。

## 下载并安装 MgC Agent

**步骤1** 使用在源端内网环境准备好的Linux主机，登录[迁移中心管理控制台](#)。

**步骤2** 在左侧导航栏单击“迁移中心Agent”，进入迁移中心Agent页面。

**步骤3** 在Linux版区域，单击“下载安装包”或“复制下载命令”，将MgC Agent安装程序下载到Linux主机。

**步骤4** 执行如下命令，解压MgC Agent安装包。

```
tar zxvf Edge.tar.gz
```

**步骤5** 执行如下命令，进入MgC Agent安装目录中的scripts目录。

```
cd Edge/scripts/
```

**步骤6** 执行如下命令，启动MgC Agent安装脚本。

```
./install.sh
```

**步骤7** 输入Linux本机网卡的弹性公网IP地址作为后续访问MgC Agent页面的地址。如果输入的地址不在本机拥有的IP列表中，会提示是否开放本机所拥有的任何公网IP作为访问地址。

```
Please enter the access address of the local server:1.1.1.1
The entered IP address is not in the local IP address list.
Do you want to allow access from all IP addresses?(y/n)
```

**步骤8** 当出现如下图所示提示时，表示Linux版的MgC Agent已安装完成。其中提示的端口号请以实际情况为准（通常为27080）。

```
There are some variables appended into /etc/profile, if you want to make these available in current terminal, please run command 'source /etc/profile'

Open the Edge management console by accessing https://'Available IP of local host':27080/ from the browser.
```

执行如下命令，更新环境变量。

```
source /etc/profile
```

**步骤9** 安装完成后，在安全组入方向规则中添加的Windows主机上打开浏览器，输入地址“https://[步骤7](#)输入的IP:[步骤8](#)提示的端口号”，即可访问MgC Agent的用户注册页面。例如：[步骤7](#)输入的IP为192.168.x.x，[步骤8](#)提示的端口号为27080，则MgC Agent的访问地址为：https://192.168.x.x:27080。

### 须知

如果遇到访问问题，请检查IP地址的准确性以及安装MgC Agent的Linux主机安全组入网规则配置。

----结束

## 注册用户

首次登录，需注册用户名、密码。请妥善保存用户名和密码，避免丢失。

**步骤1** 在用户注册页面，分别输入用户名、密码，确认密码无误后，单击《[隐私政策声明](#)》。

用户注册

首次登录，请配置用户名和密码。

用户名

密码

确认密码

我已阅读并同意《[隐私政策声明](#)》  
请先单击阅读说明，才能勾选同意

**步骤2** 仔细阅读《[隐私政策声明](#)》内容后，勾选“我已阅读并同意《[隐私政策声明](#)》”并单击“注册”按钮。

### 注意

完成用户注册后，建议每3-6个月更换一次密码。

----结束

## 连接迁移中心

**步骤1** 使用注册的用户名、密码登录MgC Agent控制台。

**步骤2** 在总览页面，单击页面右上角的“[连接](#)”按钮，右侧弹出[连接迁移中心](#)页面。

图 1-34 连接迁移中心



**步骤3** 在**第一步：输入云凭证**区域，选择在MgC服务创建项目的区域，输入华为云账号的AK/SK，单击“**查询项目**”按钮。系统会对输入的AK/SK进行验证，验证通过后，即可进行下一步。如果提示AK/SK错误，请参考[AK/SK验证失败，如何处理？](#)进行排查处理。

### 须知

如果是进行主机迁移，输入的AK/SK需要具备SMS服务的相关权限。需要的权限和配置方法请参考[SMS自定义策略](#)。

您可以选择是否将输入的AK/SK保存为目的端凭证（推荐您选择“是”）。

- 选择“否”，与迁移中心连接成功后，输入的AK/SK会被删除。
- 选择“是”，与迁移中心连接成功后，输入的AK/SK会加密保存在本地，在后续进行迁移时，可以作为目的端凭证发送给源端SMS-Agent。

**步骤4** 在**第二步：关联迁移中心项目**区域的项目下拉列表中，选择MgC控制台所[创建的迁移项目](#)，MgC Agent会将采集数据上报到该项目。

**步骤5** 在**第三步：预设当前迁移中心Agent名称**区域，自定义MgC Agent在MgC控制台显示的名称，单击“**连接**”按钮，确认要与迁移中心进行连接，单击“**确定**”按钮。

### ⚠ 注意

与迁移中心连接成功后，设置的MgC Agent名称不支持修改。

**步骤6** 当总览页面显示状态为“**已连接**”，代表与MgC连接成功。

----**结束**

## 添加资源凭证

支持在MgC Agent上添加凭证的资源类型包括：私有云、主机、大数据、容器。各资源类型包含的认证方式参见表1-22。

表 1-22 各资源认证方式

| 资源类型               | 认证方式                                                                                 | 说明                                                                                                                                                                                                                                                                                                         |
|--------------------|--------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 公有云                | <ul style="list-style-type: none"> <li>AK/SK</li> <li>配置文件</li> <li>ID/密钥</li> </ul> | <ul style="list-style-type: none"> <li><b>AK/SK</b>适用于添加<b>华为云、阿里云、AWS、腾讯云、七牛云和金山云</b>等云平台凭证。</li> <li><b>配置文件</b>适用于添加<b>谷歌云</b>平台凭证。配置文件需要上传<b>谷歌云服务账号凭据</b>，该凭证必须为<b>json</b>格式文件，且大小不能超过4KB。</li> <li><b>ID/密钥</b>适用于添加<b>Azure</b>平台凭证。添加凭证所需的信息获取方法请参见<a href="#">如何获取添加Azure凭证所需的信息</a>。</li> </ul> |
| 私有云                | 用户名/密码                                                                               | 输入源端私有云的登录用户名、密码。                                                                                                                                                                                                                                                                                          |
| 数据库                | 用户名/密码                                                                               | 输入数据库的登录用户名、密码。                                                                                                                                                                                                                                                                                            |
| 大数据-执行机            | 用户名/密码                                                                               | 输入大数据所部署主机的登录用户名、密码；网段限制填写网段地址，可以是单个IP地址或IP地址段。<br>例如： <ul style="list-style-type: none"> <li>单个IP地址：192.168.10.10/32</li> <li>IP地址段：192.168.52.0/24</li> <li>所有IP地址：0.0.0.0/0</li> </ul>                                                                                                                  |
| 大数据-Hive Metastore | 用户名/密钥                                                                               | 凭证文件需要同时上传core-site.xml、hivemetastore-site.xml、hive-site.xml、krb5.conf、user.keytab五个文件。获取方式参见 <a href="#">大数据-Hive Metastore凭证文件获取方法</a> 。                                                                                                                                                                 |
| 大数据-数据湖搜索 (DLI)    | AK/SK                                                                                | 输入华为云账号的AK/SK。获取方法请参考 <a href="#">如何获取AK/SK</a> 。                                                                                                                                                                                                                                                          |
| 大数据-MaxCompute     | AK/SK                                                                                | 输入源端阿里云账号的AK/SK。获取方法请参考 <a href="#">查看RAM用户的AccessKey信息</a> 。                                                                                                                                                                                                                                              |
| 大数据-Doris          | 用户名/密码                                                                               | 输入登录Doris数据库的用户名、密码。                                                                                                                                                                                                                                                                                       |

| 资源类型           | 认证方式                                                                     | 说明                                                                                                                                                                                                                                                                                                                                            |
|----------------|--------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 大数据-HBase      | 用户名/密钥                                                                   | <ul style="list-style-type: none"> <li>非安全集群需要上传core-site.xml, hdfs-site.xml, yarn-site.xml, mapred-site.xml和hbase-site.xml等5个文件。</li> <li>安全集群需要上传core-site.xml, hdfs-site.xml, yarn-site.xml, krb5.conf, user.keytab, mapred-site.xml和hbase-site.xml等7个文件。</li> </ul> 以上配置文件通常位于Hadoop和HBase安装目录的conf子目录下。                                  |
| 大数据-ClickHouse | 用户名/密码                                                                   | 输入登录ClickHouse数据库的用户名、密码。                                                                                                                                                                                                                                                                                                                     |
| Windows主机      | 用户名/密码                                                                   | 输入源端主机的登录用户名、密码。网段限制填写网段地址，可以是单个IP地址或IP地址段。<br>例如： <ul style="list-style-type: none"> <li>单个IP地址：192.168.10.10/32</li> <li>IP地址段：192.168.52.0/24</li> <li>所有IP地址：0.0.0.0/0</li> </ul>                                                                                                                                                         |
| Linux主机        | <ul style="list-style-type: none"> <li>用户名/密码</li> <li>用户名/密钥</li> </ul> | <ul style="list-style-type: none"> <li>选择用户名和密码时，输入源端主机的登录用户名、密码。</li> <li>选择用户名和密钥时，输入源端主机登录用户名、口令（密钥文件对应的密码）、上传.pem类型的密钥文件。</li> </ul> <b>须知</b><br>如果密钥文件未加密，则口令无需填写。<br>网段限制填写网段地址，可以是单个IP地址或IP地址段。例如： <ul style="list-style-type: none"> <li>单个IP地址：192.168.10.10/32</li> <li>IP地址段：192.168.52.0/24</li> <li>所有IP地址：0.0.0.0/0</li> </ul> |
| 容器             | 配置文件                                                                     | 配置文件格式必须为.json文件或.yml文件。                                                                                                                                                                                                                                                                                                                      |

**步骤1** 使用注册的用户名、密码登录MgC Agent控制台。

**步骤2** 在左侧导航树选择“云边协同 > 凭证列表”，进入凭证列表页面。

**步骤3** 单击列表上方的“创建凭证”按钮，弹出创建凭证窗口。

图 1-35 创建凭证



**步骤4** 按照窗口提示，选择资源类型后，选择认证方式并输入相应凭证，凭证名称为用户自定义，单击“确认”。

- 如果未连接迁移中心，添加的主机凭证可以用于本地主机深度采集。
- 如果连接了迁移中心，系统会自动将添加的凭证同步至迁移中心。

----结束

## 1.6.2 采集元数据

### 1.6.2.1 创建元数据连接

创建元数据来源类型的连接，获取源端大数据数据库表清单。

目前支持的连接类型，参见表1-23。

表 1-23 支持的元数据连接类型

| 类别     | 连接类型                                                                                                                                                                                                                               |
|--------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 大数据组件  | <ul style="list-style-type: none"> <li>● Hive Metastore</li> <li>● Doris</li> <li>● HBase</li> <li>● ClickHouse</li> <li>● Delta Lake (有元数据)</li> <li>● Delta Lake (无元数据)</li> <li>● Hudi (有元数据)</li> <li>● Hudi (无元数据)</li> </ul> |
| 大数据云服务 | <ul style="list-style-type: none"> <li>● MaxCompute</li> <li>● 阿里云云数据库ClickHouse</li> </ul>                                                                                                                                        |

## 须知

- 在创建Hive Metastore连接前，需要先获取连接Hive Metastore的IP地址和端口号。
- 在创建MaxCompute连接前，需要先获取MaxCompute服务所在地域的Endpoint信息。
- 在创建Doris连接前，需要先获取Doris集群部署的服务器IP地址和端口。
- 在创建HBase连接前，需要先获取HBase集群中至少一个ZooKeeper节点的IP地址和连接端口。

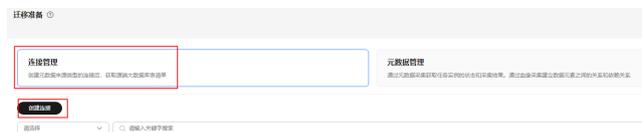
## 前提条件

已在源端元数据来源的环境中部署迁移中心Agent。

## 操作步骤

- 步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择创建的[大数据迁移项目](#)。
- 步骤2** 在左侧导航栏选择“[调研>迁移准备](#)”，进入[迁移准备](#)页面。
- 步骤3** 单击“[创建连接](#)”，右侧弹出[新增连接](#)窗口。

图 1-36 创建连接



- 步骤4** 选择连接类型，单击“[下一步](#)”，进入[详细配置](#)页面。
- 步骤5** 根据所选的连接类型，查看对应的参数配置说明，配置连接参数。
  - [Hive Metastore连接参数配置说明](#)
  - [Doris连接参数配置说明](#)
  - [HBase连接参数配置说明](#)
  - [ClickHouse组件参数配置说明](#)
  - [Delta Lake（有元数据）连接参数配置说明](#)
  - [Delta Lake（无元数据）连接参数配置说明](#)
  - [Hudi（有元数据）连接参数配置说明](#)
  - [Hudi（无元数据）连接参数配置说明](#)
  - [MaxCompute连接参数配置说明](#)
  - [阿里云云数据库ClickHouse连接参数配置说明](#)

表 1-24 Hive Metastore 连接参数配置说明

| 参数                   | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 所属源                  | 根据实际需求选择。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| 连接名称                 | 默认创建名称为“Hive-Metastore-4位随机数（字母和数字）”，也可以自定义名称。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| 迁移中心Agent            | 选择在源端环境中 <a href="#">部署迁移中心Agent</a> 。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| 启用安全连接               | 选择是否启用安全连接。 <ul style="list-style-type: none"> <li>MetaStore在非安全集群下，可以不启用安全连接。</li> <li>MetaStore在安全集群下，需要启用安全连接并选择凭证。凭证的获取以及添加方法请参考<a href="#">添加资源凭证</a>中的大数据-Hive Metastore。</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| Hive版本               | 选择源端Hive的实际版本。<br><b>注意</b><br>如果源端Hive为2.1.1版本，此处需要选择1.x版本。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| MetaStore节点IP        | 填写连接MetaStore的IP地址。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| MetaStore Thrift服务端口 | 填写连接MetaStore的端口，默认为9083。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| 启用元数据仓库连接            | <p>在增量数据校验中，如果通过Metastore查询，当分区数量超过30,000时，可能会因为将所有分区信息加载到内存中而导致内存溢出（OOM）。因此，使用MySQL作为元数据仓库可以有效地解决这一问题。</p> <ul style="list-style-type: none"> <li>不启用：系统通过MetaStore查询Hive表的分区信息。</li> <li>启用：配置MySQL作为元数据仓库，系统将通过MySQL查询Hive表的分区信息。启用后，需要配置如下参数： <ul style="list-style-type: none"> <li>元仓类型：目前仅支持MySQL作为元数据仓库。</li> <li>MySQL凭证：选择访问MySQL数据库的认证信息。认证信息需要添加到MgC Agent然后同步到MgC，方法请参考<a href="#">添加资源凭证</a>。</li> <li>MySQL节点IP：输入MySQL数据库服务器的IP地址。</li> <li>MySQL服务端口：输入MySQL数据库服务的端口号。</li> <li>数据库名：输入存储Hive表元数据的数据库名称。</li> </ul> </li> </ul> <p><b>说明</b><br/>请确保填写的MySQL凭证、节点IP、服务端口和数据库名与Hive实际使用的MySQL数据库相匹配，否则会导致数据校验失败。</p> |

表 1-25 Doris 连接参数配置说明

| 参数        | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|-----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 所属源       | 根据实际需求选择。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| 连接名称      | 默认创建名称为“Doris-4位随机数（字母和数字）”，也可以自定义名称。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| 迁移中心Agent | 选择在源端环境中 <a href="#">部署迁移中心Agent</a> 。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| Doris凭证   | 选择在MgC Agent添加的Doris凭证。凭证的添加方式请参考 <a href="#">添加资源凭证</a> ，类型选择“大数据-Doris”。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| 数据库地址     | 填写连接到Doris集群部署的服务器IP地址。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| 数据库端口     | 填写连接到Doris集群部署的服务器的端口。默认为3306。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| 数据库名      | 填写Doris数据库名称。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| 采集资源使用信息  | <p>可选参数。开启后，通过该连接创建的统计任务将收集大数据服务的资源使用信息，这些信息将用于MgC上的报告生成，以便于进行性能优化。</p> <p><b>须知</b></p> <p>使用该功能，需要MgC Agent添加的华为云账号AK/SK拥有MRS服务和DLI服务的只读权限（ReadOnly）。</p> <ul style="list-style-type: none"> <li>如果所选凭证与登录迁移中心 MgC的账号凭证相同，请勾选“<b>所选凭证与登录迁移中心凭证相同</b>”，系统将根据所选的区域，自动查询账号在该区域下的项目。 <ul style="list-style-type: none"> <li>区域：选择需要校验的数据所在区域/项目。</li> <li>项目：选择需要校验的数据所在项目。</li> <li>集群ID：输入需要校验的数据所在集群ID。</li> </ul> </li> <li>如果所选凭证与登录迁移中心 MgC的账号凭证不同，请手动输入如下参数信息： <ul style="list-style-type: none"> <li>区域ID：输入需要校验的数据所在区域ID。例如，区域为“华南-广州”，则输入“cn-south-1”</li> <li>项目ID：输入与区域对应的项目ID。</li> <li>集群ID：输入需要校验的数据所在集群ID。</li> </ul> </li> </ul> <p><b>说明</b></p> <ul style="list-style-type: none"> <li>区域ID和项目ID可以在“我的凭证 &gt; <a href="#">API凭证</a>”查看。</li> <li>集群ID的获取方法请参考获取<a href="#">MRS集群ID</a>。</li> </ul> |

表 1-26 HBase 连接参数配置说明

| 参数  | 配置说明      |
|-----|-----------|
| 所属源 | 根据实际需求选择。 |

| 参数            | 配置说明                                                                                                                                               |
|---------------|----------------------------------------------------------------------------------------------------------------------------------------------------|
| 连接名称          | 默认创建名称为“HBase-4位随机数（字母和数字）”，也可以自定义名称。                                                                                                              |
| 迁移中心Agent     | 选择在源端环境中 <a href="#">部署迁移中心Agent</a> 。<br><b>注意</b><br>需要在安装MgC Agent的主机上配置/etc/hosts后，HBase元数据连接才能连接成功。host文件的获取方法请参见 <a href="#">准备开发和运行环境</a> 。 |
| HBase凭证       | 选择在MgC Agent添加的HBase凭证。凭证的添加方式请参考 <a href="#">添加资源凭证</a> ，类型选择“大数据-HBase”。                                                                         |
| 安全集群          | 选择是否部署在安全集群下。                                                                                                                                      |
| Zookeeper节点IP | 填写连接到Zookeeper的IP地址。通常是部署Zookeeper的服务器公网IP地址或内网IP地址。                                                                                               |
| Zookeeper连接端口 | 填写连接到Zookeeper的端口。                                                                                                                                 |
| HBase 版本      | 选择HBase的版本。                                                                                                                                        |

表 1-27 ClickHouse 连接参数配置说明

| 参数                   | 配置说明                                                                            |
|----------------------|---------------------------------------------------------------------------------|
| 所属源                  | 根据实际需求选择。                                                                       |
| 连接名称                 | 默认创建名称为“ClickHouse-4位随机数（字母和数字）”，也可以自定义名称。                                      |
| 迁移中心Agent            | 选择在源端环境中 <a href="#">部署迁移中心Agent</a> 。                                          |
| ClickHouse凭证（可选）     | 选择在MgC Agent添加的ClickHouse凭证。凭证的添加方法请参考 <a href="#">添加资源凭证</a> 中的大数据-ClickHouse。 |
| 安全集群                 | 选择是否部署在安全集群下。                                                                   |
| ClickHouse Server IP | 填写连接到ClickHouse服务器的IP地址。通常是运行ClickHouse的主机IP地址。                                 |
| HTTP端口               | 非安全集群下，与ClickHouse 服务器通信的 HTTP 端口。默认为8123。                                      |
| HTTP SSL/TLS端口       | 安全集群下，与ClickHouse 服务器通信的HTTPS 端口。                                               |

表 1-28 Delta Lake（有元数据）连接参数配置说明

| 参数  | 配置说明          |
|-----|---------------|
| 所属源 | 仅需在源端创建元数据连接。 |

| 参数         | 配置说明                                                                                       |
|------------|--------------------------------------------------------------------------------------------|
| 连接名称       | 默认创建名称为“Delta-Lake-有元数据-4位随机数（字母和数字）”，也可以自定义名称。                                            |
| 迁移中心Agent  | 选择在源端环境中 <a href="#">部署迁移中心Agent</a> 。                                                     |
| 执行机凭证      | 选择执行机的登录凭证。凭证的添加方法请参考 <a href="#">添加资源凭证</a> 中的大数据-执行机。                                    |
| 执行机IP      | 填写连接执行机的IP地址。                                                                              |
| 执行机连接端口    | 填写连接执行机的端口。                                                                                |
| Spark客户端目录 | 填写Spark客户端的安装目录路径。                                                                         |
| 环境变量地址     | 填写环境变量文件（配置文件）的绝对路径，例如：/opt/bigdata/client/bigdata_env。                                    |
| SQL文件地址    | 填写用于存放执行作业时生成的SQL文件的文件夹地址。您需要拥有该文件夹的读写权限。<br><b>须知</b><br>迁移完成后，您需要手动清理该路径下生成的文件夹，以释放存储空间。 |

表 1-29 Delta Lake（无元数据）连接参数配置说明

| 参数         | 配置说明                                                                                       |
|------------|--------------------------------------------------------------------------------------------|
| 所属源        | 仅需在源端创建元数据连接。                                                                              |
| 连接名称       | 默认创建名称为“Delta-Lake-无元数据-4位随机数（字母和数字）”，也可以自定义名称。                                            |
| 迁移中心Agent  | 选择在源端环境中 <a href="#">部署迁移中心Agent</a> 。                                                     |
| 执行机凭证      | 选择执行机的登录凭证。凭证的添加方法请参考 <a href="#">添加资源凭证</a> 中的大数据-执行机。                                    |
| 执行机IP      | 填写连接执行机的IP地址。                                                                              |
| 执行机连接端口    | 填写连接执行机的端口。                                                                                |
| Spark客户端目录 | 填写Spark客户端的安装目录路径。                                                                         |
| 环境变量地址     | 填写环境变量文件（配置文件）的绝对路径，例如：/opt/bigdata/client/bigdata_env。                                    |
| SQL文件地址    | 填写用于存放执行作业时生成的SQL文件的文件夹地址。您需要拥有该文件夹的读写权限。<br><b>须知</b><br>迁移完成后，您需要手动清理该路径下生成的文件夹，以释放存储空间。 |

| 参数       | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 采集资源使用信息 | <p>可选参数。开启后，通过该连接创建的统计任务将收集大数据服务的资源使用信息，这些信息将用于MgC上的报告生成，以便于进行性能优化。</p> <p><b>须知</b></p> <p>使用该功能，需要MgC Agent添加的华为云账号AK/SK拥有MRS服务和DLI服务的只读权限（ReadOnly）。</p> <ul style="list-style-type: none"> <li>如果所选凭证与登录迁移中心 MgC的账号凭证相同，请勾选“<b>所选凭证与登录迁移中心凭证相同</b>”，系统将根据所选的区域，自动查询账号在该区域下的项目。 <ul style="list-style-type: none"> <li>区域：选择需要校验的数据所在区域/项目。</li> <li>项目：选择需要校验的数据所在项目。</li> <li>集群ID：输入需要校验的数据所在集群ID。</li> </ul> </li> <li>如果所选凭证与登录迁移中心 MgC的账号凭证不同，请手动输入如下参数信息： <ul style="list-style-type: none"> <li>区域ID：输入需要校验的数据所在区域ID。例如，区域为“华南-广州”，则输入“cn-south-1”</li> <li>项目ID：输入与区域对应的项目ID。</li> <li>集群ID：输入需要校验的数据所在集群ID。</li> </ul> </li> </ul> <p><b>说明</b></p> <ul style="list-style-type: none"> <li>区域ID和项目ID可以在“我的凭证 &gt; <b>API凭证</b>”查看。</li> <li>集群ID的获取方法请参考获取<b>MRS集群ID</b>。</li> </ul> |

表 1-30 Hudi（有元数据）连接参数配置说明

| 参数         | 配置说明                                                    |
|------------|---------------------------------------------------------|
| 所属源        | 仅需在源端创建元数据连接。                                           |
| 连接名称       | 默认创建名称为“Hudi-有元数据-4位随机数（字母和数字）”，也可以自定义名称。               |
| 迁移中心Agent  | 选择在源端环境中 <b>部署迁移中心Agent</b> 。                           |
| 执行机凭证      | 选择执行机的登录凭证。凭证的添加方法请参考 <b>添加资源凭证</b> 中的大数据-执行机。          |
| 执行机IP      | 填写连接执行机的IP地址。                                           |
| 执行机连接端口    | 填写连接执行机的端口。                                             |
| Spark客户端目录 | 填写Spark客户端的安装目录路径。                                      |
| 环境变量地址     | 填写环境变量文件（配置文件）的绝对路径，例如：/opt/bigdata/client/bigdata_env。 |

| 参数      | 配置说明                                                                                                   |
|---------|--------------------------------------------------------------------------------------------------------|
| SQL文件地址 | <p>填写用于存放执行作业时生成的SQL文件的文件夹地址。您需要拥有该文件夹的读写权限。</p> <p><b>须知</b><br/>迁移完成后，您需要手动清理该路径下生成的文件夹，以释放存储空间。</p> |

表 1-31 Hudi（无元数据）连接参数配置说明

| 参数         | 配置说明                                                                                                   |
|------------|--------------------------------------------------------------------------------------------------------|
| 所属源        | 仅需在源端创建元数据连接。                                                                                          |
| 连接名称       | 默认创建名称为“Hudi-无元数据-4位随机数（字母和数字）”，也可以自定义名称。                                                              |
| 迁移中心Agent  | 选择在源端环境中 <a href="#">部署迁移中心Agent</a> 。                                                                 |
| 执行机凭证      | 选择执行机的登录凭证。凭证的添加方法请参考 <a href="#">添加资源凭证</a> 中的大数据-执行机。                                                |
| 执行机IP      | 填写连接执行机的IP地址。                                                                                          |
| 执行机连接端口    | 填写连接执行机的端口。                                                                                            |
| Spark客户端目录 | 填写Spark客户端的安装目录路径。                                                                                     |
| 环境变量地址     | 填写环境变量文件（配置文件）的绝对路径，例如：/opt/bigdata/client/bigdata_env。                                                |
| SQL文件地址    | <p>填写用于存放执行作业时生成的SQL文件的文件夹地址。您需要拥有该文件夹的读写权限。</p> <p><b>须知</b><br/>迁移完成后，您需要手动清理该路径下生成的文件夹，以释放存储空间。</p> |

| 参数       | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 采集资源使用信息 | <p>可选参数。开启后，通过该连接创建的统计任务将收集大数据服务的资源使用信息，这些信息将用于MgC上的报告生成，以便于进行性能优化。</p> <p><b>须知</b></p> <p>使用该功能，需要MgC Agent添加的华为云账号AK/SK拥有MRS服务和DLI服务的只读权限（ReadOnly）。</p> <ul style="list-style-type: none"> <li>如果所选凭证与登录迁移中心 MgC的账号凭证相同，请勾选“<b>所选凭证与登录迁移中心凭证相同</b>”，系统将根据所选的区域，自动查询账号在该区域下的项目。 <ul style="list-style-type: none"> <li>区域：选择需要校验的数据所在区域/项目。</li> <li>项目：选择需要校验的数据所在项目。</li> <li>集群ID：输入需要校验的数据所在集群ID。</li> </ul> </li> <li>如果所选凭证与登录迁移中心 MgC的账号凭证不同，请手动输入如下参数信息： <ul style="list-style-type: none"> <li>区域ID：输入需要校验的数据所在区域ID。例如，区域为“华南-广州”，则输入“cn-south-1”</li> <li>项目ID：输入与区域对应的项目ID。</li> <li>集群ID：输入需要校验的数据所在集群ID。</li> </ul> </li> </ul> <p><b>说明</b></p> <ul style="list-style-type: none"> <li>区域ID和项目ID可以在“我的凭证 &gt; <a href="#">API凭证</a>”查看。</li> <li>集群ID的获取方法请参考获取<a href="#">MRS集群ID</a>。</li> </ul> |

表 1-32 MaxCompute 连接参数配置说明

| 参数        | 配置说明                                                                         |
|-----------|------------------------------------------------------------------------------|
| 所属源       | 仅需在源端创建元数据连接。                                                                |
| 连接名称      | 默认创建名称为“MaxCompute-4位随机数（字母和数字）”，也可以自定义名称。                                   |
| 迁移中心Agent | 选择在源端环境中 <a href="#">部署迁移中心Agent</a> 。                                       |
| 阿里云凭证     | 选择源端阿里云账号的凭证。凭证的添加方式请参考 <a href="#">添加资源凭证</a> ，类型选择“大数据-MaxCompute”。        |
| 项目名称      | 输入需要采集的MaxCompute项目名称。项目名称可在MaxCompute侧控制台获取。                                |
| Endpoint  | 输入MaxCompute服务所在地域的Endpoint信息。各地域Endpoint信息请参考 <a href="#">Endpoint对照表</a> 。 |

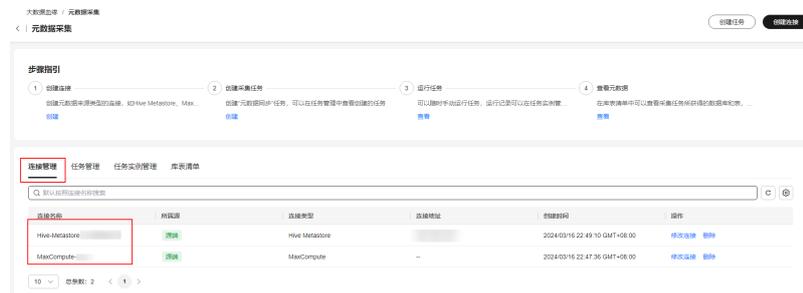
表 1-33 阿里云云数据库 ClickHouse 连接参数配置说明

| 参数           | 配置说明                                                                            |
|--------------|---------------------------------------------------------------------------------|
| 所属源          | 仅需在源端创建元数据连接。                                                                   |
| 连接名称         | 默认创建名称为“阿里云云数据库ClickHouse-4位随机数（字母和数字）”，也可以自定义名称。                               |
| 迁移中心Agent    | 选择在源端环境中 <b>部署迁移中心Agent</b> 。                                                   |
| ClickHouse凭证 | 选择在MgC Agent添加的源端阿里云云数据库ClickHouse凭证。凭证的添加方法请参考 <b>添加资源凭证</b> 中的大数据-ClickHouse。 |
| 数据库链接        | 填写连接源端ClickHouse集群的外网地址。可以在 <b>集群信息</b> 页面，查看外网地址。                              |

**步骤6** 配置完成后，单击“**测试连接**”按钮，系统会校验配置信息并与源端建立连接，测试通过代表连接成功。

**步骤7** 测试连接通过后，单击“**确认**”按钮，创建元数据连接完成。

**步骤8** 单击“**元数据连接**”，进入**元数据采集**页面。在**连接管理**页签，可以查看已创建的元数据连接列表以及基本信息。单击操作列的“**修改连接**”，可以修改连接参数配置。



----结束

### 1.6.2.2 创建并运行元数据同步任务

利用元数据连接，创建元数据同步任务，将源端数据库的库、表、字段信息同步至迁移中心。

支持创建元数据同步任务的连接类型如下：

- **Hive Metastore**连接
- **Doris**连接
- **HBase**连接
- **ClickHouse**连接
- **Hudi**（有元数据）连接
- **Hudi**（无元数据）连接
- **MaxCompute**连接

- 阿里云数据库ClickHouse连接

您可以通过手动添加、导入模板等方式添加元数据库表信息，方法请参见[查看元数据](#)。

## 前提条件

已[创建元数据连接](#)。

## 操作步骤

**步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择创建的[大数据迁移项目](#)。

**步骤2** 在左侧导航栏选择“[调研>迁移准备](#)”，进入[迁移准备](#)页面。

**步骤3** 选择[元数据管理](#)，单击[创建元数据同步任务](#)按钮，右侧弹出[创建任务-元数据同步](#)窗口。

图 1-37 创建元数据同步任务



**步骤4** 根据[表1-34](#)，配置元数据同步任务参数。

表 1-34 元数据同步任务参数配置说明

| 参数    | 配置说明                                                                                                  |
|-------|-------------------------------------------------------------------------------------------------------|
| 任务名称  | 默认创建名称为“元数据同步任务-4位随机数（字母和数字）”，也可以自定义名称。                                                               |
| 元数据连接 | 根据实际需求，选择 <a href="#">已创建的元数据连接</a> 类型。                                                               |
| 数据库   | 填写需要进行元数据同步的数据库名称。<br><b>须知</b><br>元数据连接为MaxCompute类型时，该参数为必填参数，不能为空。其他元数据连接类型，该参数为可选参数，不填则代表同步全部数据库。 |
| 并发线程数 | 设置采集线程并发数。默认为3，1≤取值范围≤10。增加线程并发数，可以提高同步效率，但会占用更多源端连接和MgC Agent所安装主机的计算资源。                             |

**步骤5** 参数配置完成后，单击“[确认](#)”按钮，创建元数据同步任务完成。

**步骤6** 在[任务管理](#)页签，可以查看已创建的元数据同步任务列表以及参数配置信息。单击操作列的“[更多>修改](#)”，可以修改任务参数配置。

图 1-38 元数据同步任务管理



- 步骤7** 单击操作列的“**运行任务**”，即可触发一次任务执行。每执行一次任务，会生成一条任务实例。
- 步骤8** 单击操作列的“**查看实例**”，进入**任务实例管理**页签。可以查看该任务的运行记录，以及每条任务实例的状态和采集结果。当元数据同步任务的实例状态为“**已完成**”且采集结果有数值时，在**库表清单**页签可以查看采集到的元数据列表。

图 1-39 任务实例管理



- 步骤9** 元数据采集完成后，在元数据列表单击血缘列的“**去采集**”，可以**创建血缘采集任务**。

图 1-40 库表清单



----结束

### 1.6.2.3 创建并运行数据湖元数据同步任务

利用创建的**Delta Lake元数据连接**，将源端数据库的库、表、字段信息同步至迁移中心。仅**Delta Lake（有元数据）**支持创建数据湖元数据同步任务。

**Delta Lake（无元数据）**只能通过手动添加、导入模板等方式添加元数据库表信息，方法请参见**查看元数据**。

#### 前提条件

已**创建Delta Lake（有元数据）元数据连接**。

#### 操作步骤

- 步骤1** 登录**迁移中心管理控制台**。在左侧导航栏的项目下拉列表中选择创建的**大数据迁移项目**。
- 步骤2** 在左侧导航栏选择“**调研>迁移准备**”，进入**迁移准备**页面。
- 步骤3** 选择**元数据管理**，单击**创建数据湖元数据同步任务**按钮，右侧弹出**创建任务-数据湖元数据同步**窗口。

图 1-41 创建数据湖元数据同步任务



- 步骤4** 根据**表1-35**，配置数据湖元数据同步任务参数。

表 1-35 元数据同步任务参数配置说明

| 参数    | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|-------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 任务名称  | 默认创建名称为“数据湖元数据同步任务-4位随机数（字母和数字）”，也可以自定义名称。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| 元数据连接 | 选择 <b>已创建的Delta-Lake（有元数据）元数据连接</b> 。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| 数据库   | 填写需要进行元数据同步的数据库名称。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| 并发线程数 | 设置采集线程并发数。默认为3，1≤取值范围≤10。增加线程并发数，可以提高同步效率，但会占用更多源端连接和迁移中心 Agent（MgC Agent）所安装主机的计算资源。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| 自定义参数 | <p>可以通过自定义参数设置同步的特定表或分区、过滤条件等。</p> <ul style="list-style-type: none"> <li>● 如果需要同步的是阿里云 EMR环境中的元数据，请添加如下参数： <ul style="list-style-type: none"> <li>- 参数名：conf</li> <li>- 参数值：spark.sql.catalogImplementation=hive</li> </ul> </li> <li>● 如果源端是阿里云 EMR Delta 2.2版本并且是通过Delta 2.3版本依赖访问时，请添加如下参数： <ul style="list-style-type: none"> <li>- 参数名：master</li> <li>- 参数值：local</li> </ul> </li> <li>● 如果源端是阿里云 EMR Delta 2.1.0版本并且配置了Spark 2.4.8来处理Delta Lake数据，请添加如下参数： <ul style="list-style-type: none"> <li>- 参数名：mgc.delta.spark.version</li> <li>- 参数值：2</li> </ul> </li> <li>● 如果源端是阿里云 EMR并且配置了Spark 3来处理Delta Lake数据，请添加如下参数： <ul style="list-style-type: none"> <li>- 参数名：jars</li> <li>- 参数值：'/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-core_2.12-*.jar,/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-storage-*.jar'</li> </ul> </li> </ul> <p><b>注意</b><br/>参数值需要根据实际环境目录和Delta Lake 的版本进行替换。</p> |

**步骤5** 参数配置完成后，单击“**确认**”按钮，创建数据湖元数据同步任务完成。

**步骤6** 在**任务管理**页签，可以查看已创建的数据湖元数据同步任务列表以及参数配置信息。单击操作列的“**更多>修改**”，可以修改任务参数配置。

**步骤7** 单击操作列的“**运行任务**”，即可触发一次任务执行。每执行一次任务，会生成一条任务实例。

**步骤8** 单击操作列的“**查看实例**”，进入**任务实例管理**页签。可以查看该任务的运行记录，以及每条任务实例的状态和采集结果。当数据湖元数据同步任务的实例状态为“**已完成**”且采集结果有数值时，在**库表清单**页签可以查看采集到的元数据列表。

**步骤9** 数据湖元数据采集完成后，在元数据列表单击血缘列的“去采集”，可以[创建血缘采集任务](#)。

----结束

### 1.6.2.4 查看元数据

您可以通过创建采集任务、手动添加、导入模板等方式获取元数据库表清单。

#### 通过采集任务获取元数据

创建采集任务的方法请参考[创建并运行元数据同步任务](#)。当元数据同步任务的实例状态为“已完成”且采集结果有数值时，在[库表清单](#)页签可以查看采集到的元数据列表。

#### 手动添加数据表

**步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择创建的[大数据迁移项目](#)。

**步骤2** 在左侧导航栏选择“[调研>迁移准备](#)”，进入[迁移准备](#)页面。

**步骤3** 选择[元数据管理](#)，单击[库表清单](#)页签。

**步骤4** 单击“[新增数据表](#)”按钮，弹出[新增数据表](#)窗口。

**步骤5** 根据[表1-36](#)，配置参数。

表 1-36 新增数据表参数说明

| 参数                                    | 说明                                      |
|---------------------------------------|-----------------------------------------|
| 元数据连接                                 | 根据实际需求，选择 <a href="#">已创建的元数据连接</a> 类型。 |
| 数据库名                                  | 用户自定义。                                  |
| 数据表名                                  | 用户自定义。                                  |
| 源表路径（可选）<br>元数据连接为Delta时，为必填项。        | 输入源端数据表在数据库中的存放路径。                      |
| 目标表路径（可选）<br>元数据连接为Delta（无元数据）时，为必填项。 | 输入目的端数据表在数据库中的存放路径。                     |
| 分区信息（可选）                              | 用户自定义。                                  |
| 是否分区表（可选）                             | 根据实际需求，选择该表是否为分区表。                      |
| 业务负责人(可选)                             | 可以指定数据表所属业务的责任人。                        |
| 分析责任人（可选）                             | 可以指定数据表的分析责任人。                          |

| 参数     | 说明                            |
|--------|-------------------------------|
| 跨源表    | 根据实际需求设置。设为跨源表后不再显示该表一致性校验结果。 |
| 说明（可选） | 您可以输入用于数据表的说明。                |

**步骤6** 参数配置完成后，单击“**确认**”按钮，完成数据表新增。在库表清单列表中可以查看新增的数据表。

----结束

## 批量导入数据表

**步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择创建的[大数据迁移项目](#)。

**步骤2** 在左侧导航栏选择“**调研>迁移准备**”，进入**迁移准备**页面。

**步骤3** 选择**元数据管理**，单击**库表清单**页签。

**步骤4** 单击“**导入数据表**”按钮，弹出**导入数据表**窗口。

**步骤5** 根据实际需求，选择**已创建的元数据连接**类型。

**步骤6** 单击“**模板下载**”，将导入模板下载到本地。打开导入模板，将需要添加的数据表对应参数填写到模板内。

### 须知

- 每个模板最多支持导入10,000张表。
- 待导入的表必须属于同一个元数据来源。
- 元数据连接为**Delta（有元数据）**连接时，模板中的**source\_path**参数为必填项。
- 元数据连接为**Delta（无元数据）**连接时，模板中的**source\_path**和**target\_path**参数为必填项。

**步骤7** 模板参数填写完成并保存后，单击导入窗口的“**添加文件**”按钮，将填写后的模板文件上传。

**步骤8** 上传成功后，单击“**确认**”按钮，系统开始导入，并且会在**任务实例管理**页签中的任务实例列表中，新增一条导入数据表的任务实例记录。您可以通过任务实例查看导入状态。

----结束

## 1.6.3 血缘采集

### 1.6.3.1 血缘采集概述

通过血缘采集，可以建立起数据元素之间的关系和依赖关系，从而形成一个清晰的数据血缘图。

目前支持血缘采集的文件类型以及采集须知，参见下表。

| 文件类型       | 采集须知                                                        |
|------------|-------------------------------------------------------------|
| DataArts作业 | 需要提前 <a href="#">导出DataArts作业</a> 到华为云OBS桶，并将桶的读写权限委托给迁移中心。 |
| SQL作业      | 需要提前导出SQL作业，并上传到华为云OBS桶，再将桶的读写权限委托给迁移中心。                    |
| 血缘模板       | 需要下载血缘模板，并填写相应数据。                                           |

### 1.6.3.2 导出作业

本节介绍导出华为云DataArts作业和阿里云DataWorks作业的方法。

#### 导出华为云 DataArts 作业

**步骤1** 登录DataArts Studio控制台。

**步骤2** 选择实例，单击“**进入控制台**”，选择对应工作空间。

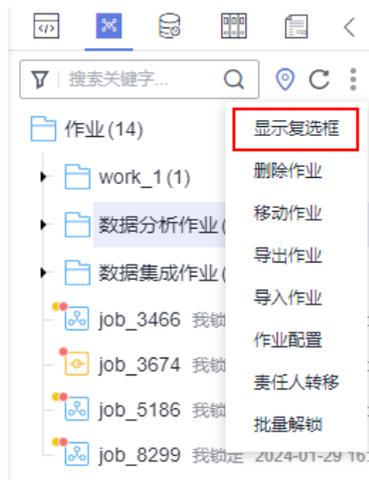
[已开通实例](#) [如何进行准备工作?](#) [快速入门](#) [免费赋能课](#)



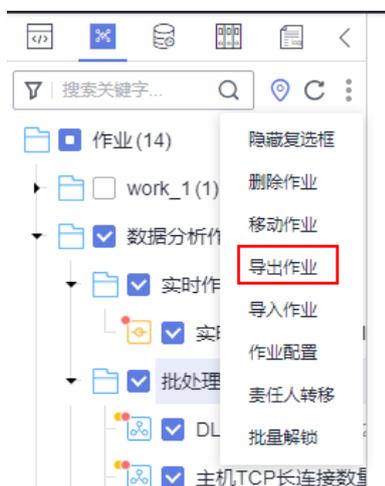
**步骤3** 单击“**数据开发**”，进入数据开发页面。



**步骤4** 单击作业目录中的 ，选择“**显示复选框**”。



**步骤5** 勾选需要导出的作业，单击  > 导出作业，弹出的**导出作业**窗口。



**步骤6** 导出方式选择“**导出到OBS**”，选择作业保存目录。

**步骤7** “导出范围”和“导出状态”根据实际需求选择。

### 须知

若待导出的作业未提交过任何版本，“导出状态”建议选择“**开发态**”，如果选择了“**提交态**”导出时会跳过该作业。

**步骤8** 单击“**确定**”，导出完成后，前往**步骤6**选择的OBS目录获取导出结果。

----结束

## 导出阿里云 DataWorks 作业

**步骤1** 登录阿里云DataWorks控制台，单击左侧导航栏的“**数据建模与开发 > 数据开发**”，在下拉框中选择对应工作空间后单击“**进入数据开发**”。

**步骤2** 单击左上方的  图标，选择全部产品 > 更多 > 迁移助手，进入迁移助手首页。

**步骤3** 参考[创建和查看DataWorks导出任务](#)，创建全量导出任务。

----结束

### 1.6.3.3 创建血缘采集任务

#### 前提条件

已完成[元数据采集](#)。

#### 操作步骤

**步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择创建的[大数据迁移项目](#)。

**步骤2** 在左侧导航栏选择“调研>迁移准备”，进入迁移准备页面。

**步骤3** 选择[元数据管理](#)，在[大数据血缘分析](#)页签，单击“创建血缘采集任务”，弹出创建血缘采集任务窗口。

图 1-42 创建血缘采集任务



**步骤4** 根据实际作业类型，选择文件类型并配置对应参数。

| 文件类型 | 参数 | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|------|----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 血缘模板 | 文件 | <p>将血缘模板下载到本地，并填写模板内的参数。其中，如下参数值为必填字段。</p> <ul style="list-style-type: none"> <li>● 目标数据库(TargetDataset)</li> <li>● 目标表(TargetTable)</li> <li>● 目标连接名称(TargetConnectionName)</li> <li>● 目标组件类型(TargetComponentType)</li> <li>● 上游数据库(SourceDataset)</li> <li>● 上游表(SourceTable)</li> <li>● 上游连接名称(SourceConnectionName)</li> <li>● 上游组件类型(SourceComponentType)</li> <li>● 作业ID(JobId)</li> </ul> <p><b>须知</b></p> <ul style="list-style-type: none"> <li>● 模板中<b>目标组件类型</b>和<b>上游组件类型</b>支持的类型包括 <b>HiveSQL</b>、<b>MaxCompute</b>。</li> <li>● 模板中单元格不允许存在公式，否则会导致解析失败。</li> </ul> <p>模板参数填写完成并保存后，单击窗口的“<b>添加文件</b>”，将填写后的模板文件上传至迁移中心。</p> <p><b>注意</b><br/>文件大小不能超过100 MB。</p> |

- **选择血缘模板。**
  - a. 单击“**模板下载**”，将血缘模板下载到本地。
  - b. 打开血缘模板，填写模板内的参数。其中，如下参数值为必填字段。
    - 目标数据库(TargetDataset)
    - 目标表(TargetTable)
    - 目标连接名称(TargetConnectionName)
    - 目标组件类型(TargetComponentType)
    - 上游数据库(SourceDataset)
    - 上游表(SourceTable)
    - 上游连接名称(SourceConnectionName)
    - 上游组件类型(SourceComponentType)
    - 作业ID(JobId)

### 须知

- 模板中目标组件类型和上游组件类型支持的类型包括HiveSQL、MaxCompute。
  - 模板中单元格不允许存在公式，否则会导致解析失败。
- c. 模板参数填写完成并保存后，单击窗口的“添加文件”，将填写后的模板文件上传至迁移中心。

### 注意

文件大小不能超过100 MB。

**步骤5** 单击“确认”，创建血缘采集任务完成。系统自动开始进行血缘采集。

**步骤6** 单击“查看任务”，进入血缘采集页面，可以在任务列表中查看采集任务状态。

**步骤7** 当任务状态为“已完成”时，单击页面右上角的“查看血缘”，可以查看[血缘关系图](#)。

----结束

## 1.6.3.4 查看血缘关系

### 前提条件

已[创建血缘采集任务](#)，且状态为“已完成”。

### 操作步骤

**步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择创建的[大数据迁移项目](#)。

**步骤2** 在左侧导航栏选择“调研>迁移准备”，进入迁移准备页面。

**步骤3** 选择元数据管理，在大数据血缘分析页签，单击“查看血缘”，进入血缘图页面。

图 1-43 查看血缘



**步骤4** 通过数据库名、表名筛选需要查看的血缘图。如果进行了大数据校验，还可以通过校验批次进行筛选。

----结束

## 1.7 迁移方案设计

## 1.7.1 绑定目的端

仅支持主机类型资源绑定目的端。如果您已经在华为云创建了目的端主机，可以将采集的源端主机与目的端主机进行绑定，绑定后，迁移 workflows 将使用所绑定的目的端作为迁移的目的端。

绑定目的端的主机，在创建迁移 workflows 前，无需进行[评估推荐](#)。

### 前提条件

已完成[资源采集](#)。

### 注意事项

绑定的目的端主机有以下要求：

- 选择绑定的目的端必须关机，如果需要绑定的目的端处于开机状态，请关机后再绑定。
- 迁移 workflows 将使用所绑定的目的端作为迁移的目的端，迁移过程中该目的端主机会被格式化，然后将源端数据迁移到该主机。
- 如果后续主机迁移 workflows 是公网迁移，所绑定目的端主机需要能访问公网。
- 选择绑定的目的端主机所在区域需要与源端主机[关联的应用](#)处于同一区域。

### 操作步骤

- 步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择创建的[应用迁移项目](#)。
- 步骤2** 单击左侧导航栏的“[设计>迁移方案设计](#)”，进入[迁移方案设计](#)页面。
- 步骤3** 单击目的端规格配置框中的“[目的端配置表](#)”，进入[目的端配置表](#)页面。
- 步骤4** 在[资源配置明细](#)的主机页签，单击主机资源绑定目的端列的“[去绑定](#)”，弹出绑定目的端窗口。
- 步骤5** 选择源端主机关联的[应用](#)所在区域以及区域下的项目，在所选项目下，选择要绑定的目的端主机。支持目的端磁盘扩容，即可以根据源端主机磁盘的实际使用量，选择小于源端磁盘容量和少于源端磁盘数量的目的端主机。

#### 须知

- 当前仅支持Linux主机磁盘扩容，需确保扩容后容量大于源端实际使用量。
- 跨可用区迁移场景仅支持扩容，即使配置扩容，系统也会自动创建一个与源端磁盘容量相同的目的端磁盘。

- 步骤6** 单击“[确认](#)”。目的端绑定成功后，该主机资源绑定目的端列的状态会变为“[已绑定](#)”，单击“[详情](#)”可以查看绑定的目的端规格。并且系统会自动判断所绑定的目的端是否进行了磁盘扩容，在“[磁盘扩容](#)”列，“是”代表已扩容；“否”代表未扩容。



----结束

## 解除目的端绑定

**步骤1** 在资源配置明细的主机页签，单击已绑定目的端的主机资源操作列的“解除目的端绑定”，弹出解除已有目的端窗口。

**步骤2** 解绑操作无法恢复，确认解除绑定后，单击“是”，完成目的端解绑。

----结束

## 1.7.2 规格评估

规格评估是根据源端资源规格、性能和应用场景以及特定的成本、可用性、性能、安全合规等需求，为您推荐最合适的华为云对应资源规格，通过规格评估确定迁移目的端。

当前支持规格评估的资源类型为：**主机、数据库、对象存储、容器、中间件**。

### 使用须知

- 已**绑定目的端**的主机资源，无需进行规格推荐，可直接创建迁移工作流。
- 规格评估结果报告有效期为3个月。

### 前提条件

已完成**资源采集**，并将采集的资源关联到**应用**。

### 操作步骤

**步骤1** 登录**迁移中心管理控制台**。在左侧导航栏的项目下拉列表中选择创建的**应用迁移项目**。

**步骤2** 单击左侧导航栏的“设计>迁移方案设计”，进入**迁移方案设计**页面。

在迁移方案设计页面，可以查看当前项目下的**资源总数**、资源的**目的端规格配置情况**以及当前项目下所创建的**应用列表**。

**步骤3** 单击**目的端规格配置框**中的“规格评估”按钮，进入**评估单应用**页面。

**步骤4** 在**选择应用**的下拉列表中，选择需要进行资源规格评估的应用。

**步骤5** 在**选择资源**区域，勾选应用内，需要评估的资源。

**步骤6** 根据**表1-37**，配置评估策略。

表 1-37 评估策略参数说明

| 参数    | 说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|-------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 购买区域  | 选择在华为云购买资源的区域。推荐您选择靠近目标用户的区域，可以减少网络时延，提高访问速度。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| 推荐策略  | <ul style="list-style-type: none"> <li>● 匹配源端规格<br/>根据源端规格推荐最合适的华为云对应资源规格，适合云上源端规格推荐。</li> <li>● 匹配业务场景<br/>以源端业务场景为基础，结合华为云最佳实践进行资源规格推荐。</li> <li>● 跨可用区迁移<br/>仅适用于华为云弹性云服务可用区下线迁移，该策略只会为应用中的<b>主机</b>资源推荐规格。选择该策略同时需要选择<b>目的可用区</b>。<br/>详细的推荐逻辑说明参见<a href="#">主机评估策略推荐逻辑说明</a>。</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                              |
| 优先级设置 | <ul style="list-style-type: none"> <li>● 性能优先<br/>推荐的目的端规格要求性能最佳。</li> <li>● 价格优先<br/>推荐的目的端规格在满足要求的前提下尽量节约成本。</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| 偏好设置  | <p>主机资源，可以进行以下偏好设置，推荐目的端时优先推荐满足偏好设置要求的主机。详细的推荐逻辑说明参见<a href="#">主机评估策略推荐逻辑说明</a>。</p> <ul style="list-style-type: none"> <li>● 虚拟机类型偏好（可选）<br/>选择期望的目的端主机类型，支持多选。</li> <li>● 虚拟机规格族偏好（可选）<br/>根据选择的主机类型，罗列出包含的规格族类型，选择期望的目的端主机规格族，支持多选。<br/><b>须知</b><br/>勾选“仅显示适合专属主机规格族”，则只能选择在购买区域所选Region下专属主机所包含的规格，并且虚拟机类型偏好将无法选择。</li> <li>● 系统盘规格偏好（可选）<br/>选择期望的目的端系统盘规格类型。</li> <li>● 数据盘规格偏好（可选）<br/>选择期望的目的端数据盘规格类型。</li> <li>● 虚拟机规格调整<br/>选择主机性能规格的推荐策略。 <ul style="list-style-type: none"> <li>- 选择“<b>不调整</b>”，会推荐与源端主机性能相同或最接近的CPU核数、内存大小。</li> <li>- 选择“<b>按性能负载调整</b>”，需要先完成<a href="#">源端主机性能采集</a>，获取源端主机性能负载数据，再通过设置数据处理规则，推荐满足条件的CPU核数、内存大小。</li> </ul> <b>须知</b><br/>采集的性能数据越多，推荐的规格越准确，建议主机性能采集时长不少于7天。</li> </ul> |

| 参数 | 说明                                                 |
|----|----------------------------------------------------|
|    | 容器资源可以设置“集群类型”、“集群版本”、“容器网络模型”等规格配置，推荐满足规格配置的集群资源。 |

**步骤7** 配置完成后，单击“**创建评估**”，主机规格评估创建成功。规格评估完成后，可以**查看评估推荐结果**和**主机性能数据**，其中目的端规格配置即为该资源的迁移目的端。

**步骤8** （可选）您还可以进行如下操作：

- **更改目的端规格**：根据实际需求，可以调整目的端主机规格和磁盘规格，并支持磁盘缩容。
- **绑定目的端**：根据评估结果，如果您在华为云上已有满足要求的主机，可以选择并绑定该主机作为目的端主机。
- **绑定专属主机**：如果您需要将源端主机迁移到华为云专属主机 DeH上，可以选择绑定已有专属主机。

----结束

## 查看评估推荐结果

在**迁移方案设计**页面的**应用列表**，单击操作列的“**查看资源配置**”，进入目的端配置表页面。

在**资源配置明细**区域，可以查看应用内各类型资源的评估推荐结果，其中**目的端规格配置**是根据源端规格配置和评估策略推荐的华为云对应规格配置，**成本预估**是华为云推荐规格所需费用。



## 查看主机性能数据

在**资源配置明细**的主机列表中，可以查看每台主机最后采集日期7天/30天内的**CPU负载**和**内存负载**平均值。单击“**性能分析**”，可以查看所有主机的性能分布统计。



## 更改目的端规格

**步骤1** 在**资源配置明细**区域，单击主机资源操作列的“**更改目的端规格**”，右侧弹出**资源详情**窗口。

**步骤2** 单击主机规格后的“更改规格”，可以修改目的端主机规格和镜像。

### 规格配置



**步骤3** 单击磁盘对应的目的端规格处的“更改规格”，可以修改目的端磁盘的存储类型和资源总用量（磁盘容量）。Linux主机的磁盘容量可以根据源端主机磁盘的实际使用量，进行缩容，即小于源端磁盘容量。更改磁盘规格后，系统会自动判断是否进行了磁盘缩容，在主机资源配置明细列表中的“磁盘缩容”列，“是”代表已缩容；“否”代表未缩容。

### 须知

- 系统盘容量取值范围为：40 G~1024 G
- 数据盘容量取值范围为：10 G~32768 G
- 当前仅支持Linux主机磁盘缩容，需确保缩容后容量大于源端实际使用量。
- 跨可用区迁移场景仅支持扩容，即使配置缩容，系统也会自动创建一个与源端磁盘容量相同的目的端磁盘。

| ^ 磁盘 (1) |              |                                      |         |         |
|----------|--------------|--------------------------------------|---------|---------|
| 是否缩容 否   |              |                                      |         |         |
| 资源类型     | 源端规格         | 目的端规格                                | 包周期价格/月 | 按需价格/小时 |
| 系统盘      | SATA   40 GB | 普通IO   40 GB<br><a href="#">更改规格</a> | ¥12.00  | ¥0.02   |

---结束

## 绑定专属主机

绑定专属主机的作用：

- 在购买资源时，系统会自动将主机购买在您绑定的专属主机上。
- 完成规格评估后，直接创建主机迁移工作流时，系统会在创建目的端主机的那一步自动将主机购买在您绑定的专属主机上。

- 步骤1** 在**资源配置明细**区域，选择主机页签，勾选需要绑定同一专属主机的主机。
- 步骤2** 单击列表上方的“**专属主机配置 > 绑定专属**”，弹出绑定专属主机窗口。

图 1-44 绑定专属主机



- 步骤3** 在下拉列表中选择要绑定的目的端专属主机，单击“**确认**”。绑定成功后，主机资源的**专属主机**列状态会变为“**已绑定**”，单击“**详情**”可以查看绑定的专属主机规格。

----结束

### 1.7.3 购买资源

目的端规格配置完成后，无需切换服务，可以直接在迁移中心批量购买目的端资源。当前仅支持购买主机资源。主机迁移工作流支持自动购买主机资源，通过“**购买列表**”批量购买资源适用于部分主机需要重新部署的场景。通过购买列表成功购买的主机，将自动绑定为迁移目的端，可以使用主机工作流继续迁移。

#### 前提条件

已对源端资源完成**规格评估**。

#### 操作步骤

- 步骤1** 登录**迁移中心管理控制台**。在左侧导航栏的项目下拉列表中选择创建的**应用迁移项目**。
- 步骤2** 单击左侧导航栏的“**设计>迁移方案设计**”，进入**迁移方案设计**页面。
- 步骤3** 单击目的端规格配置框中的“**目的端配置表**”按钮，进入**目的端配置表**页面。
- 步骤4** 在**资源配置明细**区域的主机页签，单击已完成规格评估主机购买状态列的“**添加至购买列表**”，购买状态变为“**待购买**”。
- 步骤5** 添加完成后，在页面上方的**购买列表**框中，单击“**查看列表**”按钮，进入**购买列表**页面。
- 步骤6** 在**操作流程**区域，单击“**查看我的模板**”，右侧弹出“**我的模板**”窗口，单击“**创建模板**”按钮，进入**创建配置模板**窗口。
- 步骤7** 在**基础信息**区域，选择模板类型（当前仅支持主机），自定义模板名称；在**配置信息**区域，根据**表1-38**，配置模板参数。

表 1-38 模板参数配置说明

| 参数    | 配置说明                                                                                                                                  |
|-------|---------------------------------------------------------------------------------------------------------------------------------------|
| 区域    | 选择目的端区域。                                                                                                                              |
| 项目    | 选择目的端区域项目。                                                                                                                            |
| 可用区   | 默认随机分配，也可手动选择。                                                                                                                        |
| 虚拟私有云 | 选择目的端区域创建的虚拟私有云（VPC）。                                                                                                                 |
| 子网    | 选择虚拟私有云所包含的子网，子网网段必须在所选VPC网段范围内。                                                                                                      |
| 安全组   | 所选安全组需要满足： <ul style="list-style-type: none"> <li>Windows系统开放8899端口、8900端口和22端口。</li> <li>Linux系统开放22端口。</li> </ul>                   |
| 企业项目  | 将购买的主机加入商用生产企业项目。默认项目为default。<br>企业项目管理提供了一种按企业项目管理云资源的方式，帮助您实现以企业项目为基本单元的资源及人员的统一管理。关于创建和管理企业项目的详情，请参见《 <a href="#">企业管理用户指南</a> 》。 |

**步骤8** 参数配置完成后，单击“保存”按钮，在模板列表可以查看创建的模板。

**步骤9** 模板创建完成后，返回购买列表，单击主机购买模板列的“绑定模板”，弹出“绑定购买模板”窗口，在下拉列表中选择创建的模板，单击“确认”。

**步骤10** 为主机绑定购买模板后，单击操作列的“去购买”，当购买状态为“购买成功”时，可以看到购买资源的ID。

----结束

## 1.7.4 迁移风险评估

在资源迁移前，可以通过风险评估，提前检查可能影响迁移的风险项，包括但不限于兼容性问题、固件兼容性、规格差异等。

### 须知

目前仅支持主机资源进行迁移风险评估。

## 前提条件

已通过公网发现、内网发现或手动添加中的任意一种方法获取主机资源。

## 操作步骤

**步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择创建的[应用迁移项目](#)。

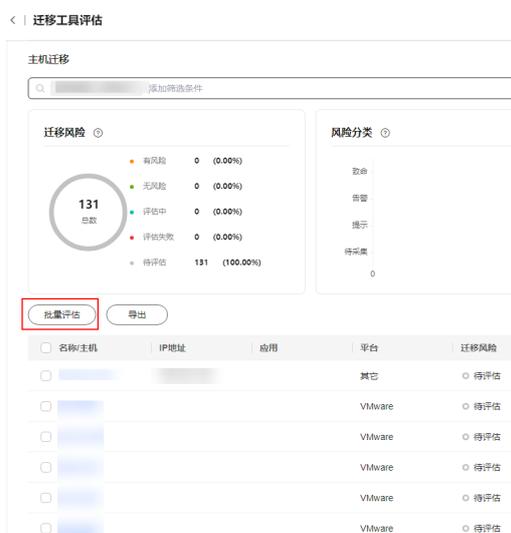
**步骤2** 单击左侧导航栏的“设计>迁移方案设计”，进入迁移方案设计页面。

**步骤3** 单击迁移工具评估框中的“查看评估表”，进入迁移工具评估页面。



**步骤4** 在主机页签，可以看到所有主机资源列表。

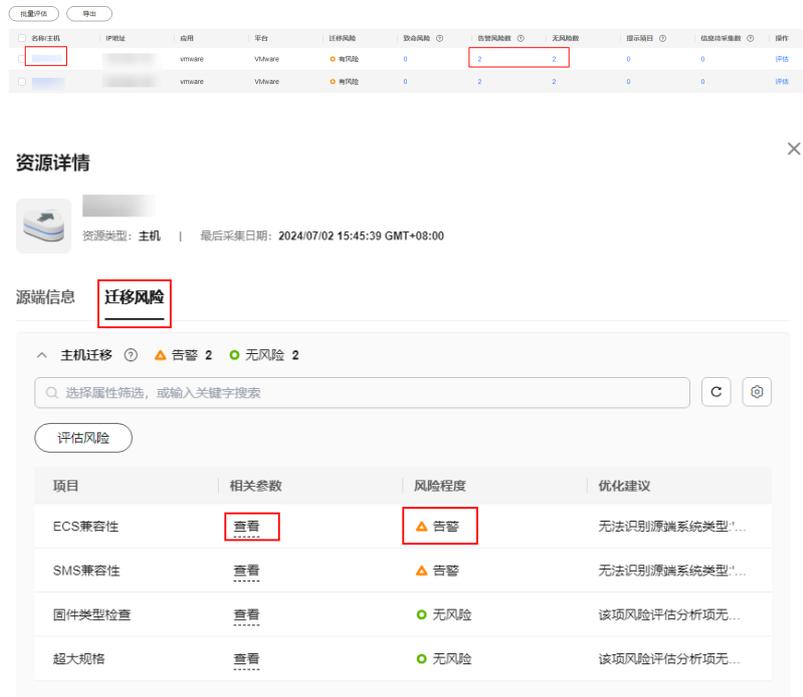
- 如果需要评估列表中所有“待评估”状态的主机，单击“批量评估”按钮，在弹出的对话框中单击“确定”，即可对所有“待评估”状态的主机进行风险评估。



- 如果仅需要评估部分主机，可以在列表中勾选需要评估的主机；也可以通过上方的搜索框，根据主机名称、IP地址、所属平台、关联的应用等条件过滤列表中的主机资源，然后勾选需要评估的主机。勾选完成后，单击“批量评估”按钮，在弹出的对话框中确认待评估主机数量无误后，单击“确认”，即可对勾选的主机进行风险评估。



**步骤5** 在评估列表中可以查看各主机的迁移风险评估状态，以及不同风险级别存在的风险项数量。单击**主机名称**或**风险项数量**，右侧弹出**资源详情**窗口，在**迁移风险**页签，可以查看评估包含的所有项目以及各项目的参数说明和风险程度。



**步骤6** 对于存在风险的项目，可以根据优化建议进行处理，处理完成后，在评估列表中，单击操作列的“**评估**”，可再次对该主机进行风险评估。



----结束

## 1.8 迁移方案配置

### 1.8.1 配置主机迁移方案

主机迁移方案是面向主机迁移 workflow 进行的一站式配置，能够实现批量评估和配置目的端主机规格。本节介绍为主机迁移配置迁移方案的方法。

#### 准备工作

- 创建**应用迁移项目**。
- 通过资源采集，完成主机资源的发现与采集。MgC提供了三种主机采集方式，您可以根据实际场景和您源端主机所在的云平台，选择适合的采集方式。
  - 源端主机位于云平台：源端为**阿里云、华为云、AWS、腾讯云、谷歌云、Azure**等云平台主机，推荐使用**公网采集**，也可以使用各场景通用的**手动添加**方式。
  - 源端主机位于线下IDC：推荐使用**内网采集**，通过**网段扫描**或**VMware扫描**，采集源端主机资源，也可以使用各场景通用的**手动添加**方式。

- 在源端内网环境中安装MgC Agent（原Edge）并注册用户，方法和要求请参见[安装Windows版本的MgC Agent](#)。安装并注册成功后，将MgC Agent与云端MgC控制台建立连接，方法请参考[连接迁移中心](#)。

## 约束与限制

- 一个方案只能创建一个迁移工作流。
- 同一台主机只能添加到一个方案内。
- 一个方案最多添加 100 台主机。

## 操作步骤

- 步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择创建的[应用迁移项目](#)。
- 步骤2** 在左侧导航栏选择“[迁移设计>迁移方案配置](#)”，进入迁移方案配置页面。
- 步骤3** 单击页面右上角的“[创建迁移方案](#)”按钮，右侧弹出迁移方案模板窗口。
- 步骤4** 单击[主机标准迁移](#)框中的“[开始配置](#)”按钮，进入主机标准迁移页面。



- 步骤5** 在[基础配置区域](#)，根据[表1-39](#)，配置相关参数。

表 1-39 基础配置参数配置说明

| 参数     | 配置说明                                           |
|--------|------------------------------------------------|
| 方案名称   | 用户自定义。                                         |
| 描述（可选） |                                                |
| 源端平台   | 选择源端主机所在的平台。<br><b>须知</b><br>同一方案下的主机需要来自同一平台。 |
| 目的端区域  | 选择需要迁移到的目的端区域。                                 |

- 步骤6** 在[待迁移主机列表区域](#)，单击“[添加主机](#)”按钮，右侧弹出[添加主机窗口](#)。



**步骤7** 勾选需要迁移的主机，单击“确认”。

### ⚠ 注意

- 同一方案下的待迁移主机，将通过同一工作流迁移，届时将使用同样迁移设置，如VPC、子网、安全组等，请自行确认该方案下主机资源范围。
- 所选的主机必须都来源于**步骤4**基础配置所选的源端平台。

**步骤8** 将待迁移的主机添加完后，单击“保存并下一步”，进入准备主机阶段。

**步骤9** 添加主机凭证。

- **手动添加**：适用于主机数量较少的场景。添加方法请参考**添加资源凭证**。
- **批量添加**：适用于主机数量较多的场景。
  - a. 单击待迁移主机列表上方的“**批量绑定凭证**”按钮，下载批量绑定凭证模板文件。打开模板文件，填写所有待迁移主机的凭证信息（用户名/密码）后保存。



- b. 将保存后的凭证模板文件导入与**MgC控制台**连接的**MgC Agent**上，系统将自动为主机绑定迁移代理和凭证。



为主机添加凭证后，您还可以单击操作列的“**修改**”，调整接入IP、接入端口、迁移代理和凭证等参数值。

**步骤10** 在待迁移主机列表中，勾选所有主机，单击列表上方的“**批量准备度检查**”按钮，系统会对待迁移的主机进行迁移准备度检查。



- 当**准备度检查**列状态会变为“**已完成**”时，代表该主机已满足迁移条件。
- 当**准备度检查**列状态提示“**缺少必要信息**”时，请单击深度采集列的“**采集**”，进行深度采集。
- 当**准备度检查**列状态提示**网络相关问题**时，请检查您填写的接入参数是否正确，确认无误后重新进行准备度检查。

**步骤11** 当所有待迁移主机的准备度检查状态均为“**已完成**”后，单击“**下一步**”，进入可选步骤“**性能采集**”。

**步骤12** （可选）主机性能采集。通过采集源端主机的实际性能负载，在后续的规格推荐中，自动调整目的端主机规格。**性能采集通过在源端抽样获取，每5分钟抽样一次，需要较长时间，请自行选择是否执行。**更多约束与限制请查看**性能采集使用须知**。如果不需要进行主机性能采集，直接单击“**下一步**”，进入“**目的端配置**”阶段。

勾选需要进行性能采集的主机，单击列表上方的“**批量性能采集 > 开始**”，启动主机性能采集。此时性能采集列的状态为“**采集中**”。

开始性能采集后，您还可以进行如下操作：

- 单击操作列的“**查看性能详情**”，可以查看采集到的主机性能数据。
- 单击列表上方的“**批量性能采集 > 停止**”，可以批量暂停性能数据采集。

**步骤13** 配置目的端。您可以根据实际情况，选择配置目的端主机的方法。

- **绑定目的端。**如果您已在华为云购买了目的端主机，可以将源端主机与目的端主机进行绑定，后续创建的迁移 workflows 将使用所绑定的主机作为迁移目的端。关于绑定的目的端主机要求请参见**绑定目的端注意事项**。

a. 在待迁移主机列表，单击绑定目的端列的“**去绑定**”，弹出**绑定目的端**窗口。



b. 选择目的端主机所在区域以及区域下的项目，在所选项目下，选择要绑定的目的端主机。支持目的端磁盘扩容，即可以根据源端主机磁盘的实际使用量，选择小于源端磁盘容量和少于源端磁盘数量的目的端主机。

## 须知

- 请确保目的端主机绑定了弹性公网IP。
  - 当前仅支持Linux主机磁盘扩容，需确保扩容后容量大于源端实际使用量。
  - 跨可用区迁移场景仅支持扩容，即使配置扩容，系统也会自动创建一个与源端磁盘容量相同的**目的端磁盘**。
- c. 单击“**确认**”。目的端绑定成功后，该主机资源**绑定目的端**列的状态会变为“**已绑定**”，单击“**详情**”可以查看绑定的目的端规格。并且系统会自动判

断所绑定的目的端是否进行了磁盘扩容，在“**磁盘扩容**”列，“是”代表已扩容；“否”代表未扩容。

- **批量规格推荐。**根据源端主机规格，系统会自动推荐在华为云与之对应的主机规格作为目的端，您还可以对推荐的主机规格进行手动调整。后续创建的工作流会按所配置规格自动创建目的端主机。**已绑定目的端的主机资源，无需进行规格推荐。**
  - a. 在待迁移主机列表，勾选需要进行规格推荐的主机，单击列表上方的“**批量规格推荐**”，右侧弹出批量规格评估窗口。
  - b. 根据**表1-40**，配置评估策略。

**表 1-40** 评估策略参数说明

| 参数    | 说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|-------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 优先级设置 | <ul style="list-style-type: none"> <li>▪ 性能优先<br/>推荐的目的端规格要求性能最佳。</li> <li>▪ 价格优先<br/>推荐的目的端规格在满足要求的前提下尽量节约成本。</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| 偏好设置  | <p>推荐目的端时优先推荐满足偏好设置要求的主机。详细的推荐逻辑说明参见<a href="#">主机评估策略推荐逻辑说明</a>。</p> <ul style="list-style-type: none"> <li>▪ 虚拟机类型偏好（可选）<br/>选择期望的目的端主机类型，支持多选。</li> <li>▪ 虚拟机规格族偏好（可选）<br/>根据选择的主机类型，罗列出包含的规格族类型，选择期望的目的端主机规格族，支持多选。</li> </ul> <p><b>须知</b><br/>勾选“仅显示适合专属主机规格族”，则只能选择在购买区域所选Region下专属主机所包含的规格，并且虚拟机类型偏好将无法选择。</p> <ul style="list-style-type: none"> <li>▪ 系统盘规格偏好（可选）<br/>选择期望的目的端系统盘规格类型。</li> <li>▪ 数据盘规格偏好（可选）<br/>选择期望的目的端数据盘规格类型。</li> <li>▪ 虚拟机规格调整<br/>选择主机性能规格的推荐策略。 <ul style="list-style-type: none"> <li>○ 选择“<b>不调整</b>”，会推荐与源端主机性能相同或最接近的CPU核数、内存大小。</li> <li>○ 选择“<b>按性能负载调整</b>”，需要先完成<a href="#">源端主机性能采集</a>，获取源端主机性能负载数据，再通过设置数据处理规则，推荐满足条件的CPU核数、内存大小。</li> </ul> </li> </ul> <p><b>须知</b><br/>采集的性能数据越多，推荐的规格越准确，建议主机性能采集时长不少于7天。</p> |

- c. 配置完成后，单击“**确认**”，系统开始进行主机规格评估。当目的端配置状态为“**成功**”时，代表规格评估完成，目的端规格即为该主机目的端规格。



- d. (可选) 更改目的端规格。您可以对推荐的目的端主机规格进行手动调整。
- 单击主机资源操作列的“**更改目的端规格**”，右侧弹出资源详情窗口。
  - 单击主机规格后的“**更改规格**”，可以修改目的端主机规格和镜像。

### 规格配置



- 单击磁盘对应的目的端规格处的“**更改规格**”，可以修改目的端磁盘的存储类型和资源总用量（磁盘容量）。Linux主机的磁盘容量可以根据源端主机磁盘的实际使用量，进行缩容，即小于源端磁盘容量。更改磁盘规格后，系统会自动判断是否进行了磁盘缩容，在主机资源配置明细列表中的“**磁盘缩容**”列，“是”代表已缩容；“否”代表未缩容。

### 须知

- 系统盘容量取值范围为：40 G~1024 G
- 数据盘容量取值范围为：10 G~32768 G
- 当前仅支持Linux主机磁盘缩容，需确保缩容后容量大于源端实际使用量。
- 跨可用区迁移场景仅支持扩容，即使配置缩容，系统也会自动创建一个与源端磁盘容量相同的目的端磁盘。

| ^ 磁盘 (1) |              |                             |         |         |
|----------|--------------|-----------------------------|---------|---------|
| 资源类型     | 源端规格         | 目的端规格                       | 包周期价格/月 | 按需价格/小时 |
| 系统盘      | SATA   40 GB | 普通IO   40 GB<br><b>更改规格</b> | ¥12.00  | ¥0.02   |

- **批量配置规格**。可以批量为待迁移主机设置同一规格的目的端主机。

### ⚠ 注意

批量配置规格为手动配置，操作之前需要您自行评估并承担相关风险。请确保您选择的规格完全符合您的系统需求和兼容性标准。

- a. 在待迁移主机列表，勾选需要迁移到同一规格的主机，单击列表上方的“批量配置规格”，右侧弹出批量配置规格窗口。
- b. 选择目的端主机规格后，单击“确认”。设置成功后主机的“目的端配置状态”会显示为“手动设置”。



- **批量设置镜像。**可以批量为待迁移主机设置同一类型的镜像。

**注意**

批量设置镜像为手动配置，需要您自行评估并承担相关风险。请确保您选择的镜像完全符合您的系统需求和兼容性标准。

- a. 在待迁移主机列表，勾选需要设置为同一种镜像类型的主机，单击列表上方的“批量设置镜像”，弹出批量设置镜像窗口。
- b. 根据实际需求，选择镜像类型和规格，单击“确认”。请确保目的端主机的系统盘容量大于镜像所需空间。

- **高级批量配置。**通过导出、导入的方式，批量进行目的端规格、磁盘、镜像设置，适合专业迁移服务人员。

- a. 在待迁移主机列表，单击列表上方“高级批量配置 > 导出目的端配置”，将列表中的所有主机信息以Excel文件格式下载到本地。



- b. 打开导出的文件，选择第一个sheet页，名为“servers”。其中蓝色列（A~G列）为源端主机信息，无需修改；黄色列（H~K列）为目的端主机信息，需要配置。

| ID    | Name    | ConnType | ConnPort | Spec | DiskSize | DiskType | Export                       | ImageType | Image                                              |
|-------|---------|----------|----------|------|----------|----------|------------------------------|-----------|----------------------------------------------------|
| 14201 | example | SSH      | 22       | 4核8G | 50GB     | SSD      | Red Hat Enterprise Linux 7.9 | public    | Public-CAD-HCC-BareTemplate-2.0.2405.1-nfs_04_0144 |

- c. 根据表1-41，配置目的端主机信息（H~K列）。

表 1-41 H~K 列参数配置说明

| 参数     | 配置说明                                                                                                                                                          |
|--------|---------------------------------------------------------------------------------------------------------------------------------------------------------------|
| flavor | <p>填入目的端主机规格名称。在第二个sheet页（名为flavors），列出了目的端区域在售的S（通用型）和C（计算增强型）系列的规格。选择合适的规格，将规格名称复制到该参数单元格。</p> <p><b>说明</b><br/>如果进行过规格评估，导出结果包含评估结果；如果未进行过规格评估，导出结果为空。</p> |

| 参数                                | 配置说明                                                                                                                                                                                                                                                                       |
|-----------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| disks(name,osDisk,type,capacity;) | 填入所有磁盘信息，多个磁盘之间以英文分号 (;) 隔开。每块磁盘必须包含 <b>名称</b> ， <b>是否为系统盘</b> ， <b>磁盘类型</b> ， <b>磁盘大小</b> 等信息，信息之间以英文逗号 (,) 隔开。磁盘大小只写数字，单位默认是GiB。<br>例如： /dev/sda,true,SAS,1024;/dev/sdb,false,SAS,2046<br><b>说明</b><br>如果进行过规格评估，导出结果包含评估结果；如果未进行过规格评估，磁盘命名为源端磁盘名称，磁盘类型为SAS，磁盘大小为源端磁盘大小。 |
| image_type                        | 填入镜像类型。<br><ul style="list-style-type: none"> <li>public (公共镜像)</li> <li>private (私人镜像)</li> <li>shared (共享镜像)</li> </ul>                                                                                                                                                  |
| image                             | 填入镜像名称。可以在第三个sheet页 (名为images)，查看目的端区域所有可用的镜像。                                                                                                                                                                                                                             |

- d. 所有主机的目的端信息配置完成后，保存Excel文件。返回MgC控制台，单击待迁移主机列表上方的“高级批量配置 > 导入目的端配置”，弹出导入目的端配置窗口。



- e. 单击窗口中的“前往OBS存放文件”，跳转到对象存储服务 OBS的桶列表页面。



- f. 选择要存放目的端配置文件 (Excel文件) 的桶，将Excel文件上传到该桶中。  
g. 上传成功后，返回MgC控制台，在导入目的端配置窗口的下拉列表中，选择存放在OBS桶中的目的端配置文件。

- h. 单击“确认”导入目的端配置文件。

#### 📖 说明

- flavor参数为空时，忽略该行。
- disk或image参数为空时，保持原有配置。
- disk数量跟源端不一致时，会报磁盘数量不一致错误。
- 磁盘类型跟flavor冲突时，会报磁盘类型错误。
- **专属主机配置**。如果您需要将源端主机迁移到华为云专属主机 DeH上，可以选择绑定已有专属主机。绑定后，主机迁移工作流会在创建目的端主机的那一步自动将主机购买在您绑定的专属主机上。
  - a. 在待迁移主机列表，勾选需要绑定专属主机的主机。
  - b. 单击列表上方的“**专属主机配置 > 绑定专属**”，弹出绑定专属主机窗口。
  - c. 在下拉列表中选择要绑定的目的端专属主机，单击“**确认**”。绑定成功后，该主机**专属主机**列的状态会变为“**已绑定**”，单击“**详情**”可以查看绑定的专属主机规格。

**步骤14** 所有待迁移主机的目的端配置完成后，根据需求选择后续操作。

- **立即创建主机迁移工作流**  
单击“**创建工作流**”按钮，进行迁移工作流参数配置。参数配置说明请参见[主机迁移工作流参数配置说明](#)。
- **仅保存迁移方案**  
单击“**确定**”，将该方案保存在迁移方案列表中。可以对保存的迁移方案进行如下操作：
  - **修改迁移方案**：单击操作列的“**设计**”，可以对方案中的配置进行修改。
  - **创建主机迁移工作流**：单击操作列的“**创建工作流**”，使用该方案创建主机迁移工作流。参数配置说明请参见[主机迁移工作流参数配置说明](#)。

---结束

## 1.8.2 配置存储批量迁移方案

迁移中心为多种资源的不同迁移方案配置模板，可以面向迁移工作流进行一站式配置，并按应用显示配置进度。本节介绍为对象存储批量迁移配置迁移方案的方法。

**对象存储批量迁移**：适用于源端有多个对象存储桶资源需要迁移的场景。如果只有单个对象存储桶资源需要迁移，请直接[创建单个对象存储迁移工作流](#)。

### 前提条件

- 已通过资源采集的公网采集，完成对象存储资源的发现与采集。
- （可选）已对待迁移的对象存储资源进行深度采集。

### 注意事项

- 创建方案时，待迁移桶列表中的资源所属的平台必须和所选的源端平台一致，且一个方案对应一个目的端区域。
- 一个源端桶资源可以关联多个方案。

- 在一个方案下，一个桶资源的目的端前缀只能有一个，但是源端前缀可以有多个。

## 操作步骤

- 步骤1** 登录**迁移中心**管理控制台。
- 步骤2** 在左侧导航栏选择“设计>迁移方案配置”，进入迁移方案配置页面。
- 步骤3** 单击页面右上角的“创建迁移方案”按钮，右侧弹出迁移方案模板窗口。
- 步骤4** 单击对象存储批量迁移框中的“开始配置”按钮，进入对象存储批量迁移页面。
- 步骤5** 在基础配置区域，根据表1-42，配置相关参数。

表 1-42 基础配置参数配置说明

| 参数     | 配置说明          |
|--------|---------------|
| 方案名称   | 用户自定义。        |
| 描述（可选） | 用户自定义。        |
| 源端平台   | 选择源端资源所在的云平台。 |
| 目的端桶区域 | 选择目的端桶所在区域。   |

- 步骤6** 在待迁移桶列表区域，单击“添加桶”按钮，右侧弹出添加桶窗口。
- 步骤7** 勾选需要迁移的桶，单击操作列的“修改”，调整桶的迁移方式，迁移方式说明参见表1-43。勾选并调整完成后，单击“确认”。

### 注意

- 所选的资源必须都来源于基础配置所选的源端平台。
- 最多添加 100 个桶。

表 1-43 迁移方式说明

| 迁移方式 | 说明                                     |
|------|----------------------------------------|
| 全量迁移 | 迁移源端桶/文件夹内的全部数据。                       |
| 列表迁移 | 通过列表文件过滤迁移，只迁移列表文件中的对象。列表文件需要存放在目的端桶中。 |
| 前缀迁移 | 通过对象名称或者名称前缀过滤迁移，只迁移与对象名称或名称前缀一致的对象。   |

- 步骤8** 在待迁移桶列表中，单击操作列的“修改”，弹出修改迁移设置窗口。
- 步骤9** 选择源端凭证，输入对象个数和容量（可选），单击“确认”，完成桶的迁移设置。

- 如果桶的迁移方式为“**列表迁移**”，需要在**列表路径**中输入列表文件的存放路径并以“/”结尾。
- 如果桶的迁移方式为“**前缀迁移**”，在“**前缀路径**”添加对象名称或者名称前缀，单击“**确定**”保存前缀。

#### 说明

输入对象前缀路径时，如果对象文件在源端桶的根目录下，则直接添加前缀；如果对象文件存在目录路径，则需要对对象前缀前添加对象文件所在的目录路径。格式为：文件夹/对象前缀

- 步骤10** 所有待迁移桶的迁移设置完成后，单击“**下一步**”，进行目的端配置。
- 步骤11** 单击操作列的“**修改**”，选择目的端凭证和目的端桶，输入目的端前缀（可选），单击“**保存**”。
- 步骤12** 所有资源的目的端配置完成后，单击“**下一步**”，进行集群评估。集群评估为可选操作，您可以跳过该步骤，直接绑定已有集群，也可以通过评估，获取集群规格推荐后，去创建集群。集群的创建方法请参考[创建集群](#)。
- 步骤13** 单击“**下一步**”，进入集群配置，绑定已有迁移集群，本方案下所有资源将通过此集群迁移。
- 步骤14** 单击“**立即绑定**”，右侧弹出**集群选择**窗口。
- 步骤15** 在集群列表中，选择需要使用的集群，单击“**确认**”。如果需要对集群进行修改，请参考[集群管理](#)。
- 步骤16** 集群配置完成后，单击“**确定**”，迁移方案创建完成。在**迁移方案配置**页面的**方案列表**，可以查看已创建的迁移方案。
- 如果需要对迁移方案配置进行调整，可以单击操作列的“**设计**”进行修改。
  - 当**设计进度**为已完成时，单击操作列的“**创建工作流**”，即可[创建对象存储批量迁移工作流](#)。

---结束

### 1.8.3 配置文件存储批量迁移方案

本节介绍为文件存储批量迁移配置迁移方案的方法。

**文件存储批量迁移**：适用于源端有多个文件存储地址内的资源需要迁移的场景。如果只有单个文件存储地址内的资源需要迁移，请直接[创建单个文件存储迁移工作流](#)。

#### 准备工作

- **创建迁移项目。**  
在MgC控制台为本次迁移创建独立的项目，方法请参见[项目管理](#)。
- **创建迁移集群。**  
通过集群可以创建Master节点、迁移节点和列举节点，确保文件存储工作流的顺利运行。创建方法请参考[创建集群](#)。
- **创建目的端弹性文件系统。**  
在目的端创建相应的弹性文件系统，用于迁移后的数据存储和管理，方法请参见[创建弹性文件系统](#)或[创建高性能文件系统](#)。

## 注意事项

- 一个方案只能创建一个迁移工作流。
- 一个方案最多添加100个文件存储。
- 同一方案下的目的端文件存储必须处于同一区域。

## 操作步骤

- 步骤1** 登录**迁移中心管理控制台**。
- 步骤2** 在左侧导航栏选择“设计>迁移方案配置”，进入**迁移方案配置**页面。
- 步骤3** 单击页面右上角的“创建迁移方案”按钮，右侧弹出**迁移方案模板**窗口。
- 步骤4** 单击**文件存储批量迁移**框中的“开始配置”按钮，进入**文件存储批量迁移**页面。
- 步骤5** 在**基础信息**区域，根据**表1-44**，配置相关参数。

表 1-44 基础信息参数配置说明

| 参数        | 配置说明           |
|-----------|----------------|
| 方案名称      | 用户自定义。         |
| 描述（可选）    | 用户自定义。         |
| 目的端文件存储区域 | 选择目的端文件存储所在区域。 |

- 步骤6** 单击“添加文件存储”按钮，在下方的列表中会新增一行源端文件存储配置信息。

图 1-45 添加源端文件存储



- 步骤7** 根据**表1-45**，配置源端文件存储信息。

表 1-45 源端文件存储参数配置说明

| 参数   | 说明                                                                                  | 须知 |
|------|-------------------------------------------------------------------------------------|----|
| 存储类型 | 支持的源端存储类型包括：<br>NAS_GFS、<br>NAS_NFS_V3_PROTOCOL、<br>NAS_NFS_V3_MOUNT，请根据<br>实际情况选择。 | -  |

| 参数       | 说明                                                                                                                                                                                                                                                                                                 | 须知                                                                                                                                                                                                                                                                                         |
|----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 迁移方式     | <p>选择迁移方式。</p> <ul style="list-style-type: none"> <li>全量迁移：迁移源端文件存储服务地址+路径内的全部数据。</li> <li>列表迁移：通过列表文件过滤迁移，只迁移列表文件中的对象。</li> </ul>                                                                                                                                                                 | -                                                                                                                                                                                                                                                                                          |
| 列表路径     | <p>迁移方式选择列表迁移时，填写列表文件所在路径。迁移方式选择全量迁移时，无需填写。</p>                                                                                                                                                                                                                                                    | <p>列表文件需要满足如下要求：</p> <ul style="list-style-type: none"> <li>列表文件大小不能超过30 MB。</li> <li>列表文件类型必须为.txt类型。</li> <li>列表文件必须是UTF-8无BOM格式编码格式。</li> <li>列表文件每行只能包含一个文件名称，且文件名称使用URL Encode编码。</li> <li>列表文件行长度不超过<math>16 \times 1024 = 16384</math>字节，否则迁移失败。</li> <li>列表文件行数不超过10000</li> </ul> |
| 文件存储服务地址 | <p>输入源端文件系统的挂载地址。</p> <ul style="list-style-type: none"> <li>存储类型为NAS_GFS时，文件存储服务地址可以填写主机名、域名、可选的子域、灵活的端口号和路径表示。</li> <li>存储类型为NAS_NFS_V3_PROTOCOL和NAS_NFS_V3_MOUNT时，格式为： <ul style="list-style-type: none"> <li>IP地址:/</li> <li>IP地址:/ xxx</li> <li>域名:/</li> <li>域名:/ xxx</li> </ul> </li> </ul> | <p>无论是全量迁移还是列表迁移，源端文件存储服务地址+路径的组合或源端文件存储服务地址+列表路径的组合都必须是唯一的不能重复，都遵循前缀匹配规则。</p> <p>例如，已有组合是地址192.168.0.1:/a和路径/abc/a，那么添加或修改另一个地址为192.168.0.1:/a和路径/abc时，会因重复而报错。要覆盖已有路径，需要先删除旧的地址和路径组合。</p>                                                                                                |
| 路径       | <p>输入待迁移文件路径，格式为：/文件夹。</p>                                                                                                                                                                                                                                                                         | <p>列表迁移时，路径填写为/即可。</p>                                                                                                                                                                                                                                                                     |

**步骤8** 配置完成后，单击操作列的“保存”。所有待迁移文件存储添加完成后，单击“下一步”，进行目的端配置。

**步骤9** 在列表中单击操作列的“修改”，根据表1-46，配置目的端文件存储信息。

表 1-46 目的端文件存储参数配置说明

| 参数       | 说明                                                                                                                                                                                                                                                                                                                                                                                     | 须知                  |
|----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|
| 源端路径     | 根据源端文件存储服务地址+路径的组合自动生成。                                                                                                                                                                                                                                                                                                                                                                | -                   |
| 存储类型     | 目前只支持 NAS_NFS_V3_MOUNT。                                                                                                                                                                                                                                                                                                                                                                | -                   |
| 文件存储服务地址 | <ul style="list-style-type: none"> <li>产品类型选择通用文件系统时，输入目的端文件系统的挂载地址。挂载地址可以在文件系统列表页面，单击<b>挂载地址</b>后面的  图标复制获取。</li> <li>产品类型选择SFS Turbo时，输入目的端文件系统的共享路径。共享路径可以在SFS Turbo文件系统列表页面，单击<b>共享路径</b>后面的  图标复制获取。</li> </ul> | 文件存储服务地址+路径的组合可以重复。 |
| 路径       | 输入目的端接收文件的路径，格式为：/文件夹。                                                                                                                                                                                                                                                                                                                                                                 | -                   |

**步骤10** 配置完成后，单击操作列的“保存”。所有资源的目的端配置完成后，单击“下一步”，进行集群评估。集群评估为可选操作，您可以跳过该步骤，直接绑定已有集群，也可以通过评估，获取集群规格推荐后，去创建集群。集群的创建方法请参考[创建集群](#)。

**步骤11** 单击“下一步”，进入集群配置，绑定已有迁移集群，本方案下所有资源将通过绑定的集群迁移。

**步骤12** 单击“立即绑定”，右侧弹出**集群选择**窗口。

**步骤13** 在集群列表中，选择需要使用的集群（要求集群状态为健康），单击“确定”。如果需要~~对~~对集群进行修改，请参考[集群管理](#)。

图 1-46 选择集群



**步骤14** 集群配置完成后，单击“确定”，迁移方案创建完成。在**迁移方案配置**页面的**方案列表**，可以查看已创建的迁移方案。

- 如果需要对迁移方案配置进行调整，可以单击操作列的“设计”进行修改。
- 当设计进度为已完成时，单击操作列的“创建工作流”，即可[创建文件存储批量迁移工作流](#)。

----结束

## 1.9 迁移集群

### 1.9.1 创建集群

迁移集群是专为存储工作流提供协作中的核心组件，通过集群可以创建和管理迁移节点、列举节点，部署和升级迁移插件，确保存储工作流顺利运行。

#### 使用须知

- 单个账号在每个区域（Region）最多只能创建2个集群。
- 为了保证迁移的稳定性和数据安全，创建的迁移集群不支持用户自行登录。如您有特殊需求必须登录，请联系技术支持人员。
- 当集群超过30天未使用时会被回收（不会删除您的资源，但是集群将变为“已废弃”状态，且不支持恢复）。
- 迁移集群所包含的资源类型以及所有节点的安全组配置详细信息请查看[资源类型和安全组配置说明](#)。

#### 费用说明

在创建迁移集群时，会产生以下费用：

- 创建Master节点、迁移节点和列举节点会产生ECS服务费用。具体计费详情请参见[ECS按需计费说明](#)或[ECS价格计算器](#)。
- 如果您使用公网迁移时，会产生NAT网关服务费用。计费详情请参见[NAT网关计费说明](#)或[NAT价格计算器](#)。
- 启用LTS服务时，会产生LTS服务费用。具体计费详情请参见[LTS计费说明](#)或[LTS价格计算器](#)。

#### 操作步骤

- 步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择创建的[应用迁移项目](#)。
- 步骤2** 在左侧导航栏选择“资源准备>迁移集群”，进入[迁移集群](#)页面。
- 步骤3** 单击页面右上角的“创建集群”按钮，进入[创建集群](#)页面。

如果是首次创建迁移集群，需要您同意华为云账号进行相关的委托授权。单击 ，可以查看委托权限所包含的细粒度权限。



**步骤4** 根据表1-47，配置参数。

**表 1-47** 集群参数配置说明

| 区域   | 参数            | 配置说明                        | 限制说明                                                                                                                                                                                                                            |
|------|---------------|-----------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 基础配置 | 集群名称          | 自定义名称。                      | 同一账户下集群名称不可重名。                                                                                                                                                                                                                  |
|      | 区域            | 选择创建集群的区域。                  | 与目的端区域保持一致。                                                                                                                                                                                                                     |
|      | 集群类型          | 选择集群使用场景。                   | 当前仅支持存储迁移场景。                                                                                                                                                                                                                    |
| 节点创建 | Master节点ECS规格 | 用于管理迁移节点和列举节点。              | 节点数量为1，无法修改。                                                                                                                                                                                                                    |
|      | 迁移节点ECS规格     | 迁移节点用于执行迁移和对比的动作，推荐规格8U16G。 | <ul style="list-style-type: none"> <li>规格选定后，后续无法更改。</li> <li>设置的节点数量需要同时满足如下要求：                             <ul style="list-style-type: none"> <li>迁移节点数+列举节点数+1 ≤ 100</li> <li>迁移节点数+列举节点数+1 ≤ 子网可用IP数</li> </ul> </li> </ul> |
|      | 列举节点ECS规格     | 列举节点用于执行列举动作，推荐规格8U16G。     |                                                                                                                                                                                                                                 |
| 网络配置 | 虚拟私有云         | 在下拉列表中选择虚拟私有云。              | -                                                                                                                                                                                                                               |
|      | 子网            | 选择与虚拟私有云网段相同且满足可用IP数的子网。    | 子网的可用IP数≥迁移节点数+列举节点数+1。                                                                                                                                                                                                         |

| 区域   | 参数         | 配置说明                                                                                                                                                                                                                                                 | 限制说明                                                                                               |
|------|------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------|
|      | 模式选择       | <ul style="list-style-type: none"> <li>公网：使用公网迁移，并选择公网网关。如果没有可用网关，选择“新建网关”，选择网关规格和EIP（EIP最多支持同时勾选20个）。</li> <li>内网：使用内网迁移，适用于同一区域内的迁移场景。</li> <li>专线：使用专线迁移，通过专线直接访问源端数据。搭建云专线的方法可参考<a href="#">云专线 DC</a>。</li> </ul>                               | -                                                                                                  |
| 高级设置 | DNS配置（可选）  | 填写DNS服务器IP地址，用于追加/etc/resolv.conf文件中的nameserver参数值。多个DNS服务器地址使用英文逗号分隔，例如：192.0.2.1,192.0.2.2                                                                                                                                                         | 最多可输入3个DNS服务器地址。                                                                                   |
|      | 域名映射配置（可选） | 添加域名到IP地址的映射，用于修改/etc/hosts文件。                                                                                                                                                                                                                       | 最多可添加500个IP地址映射。                                                                                   |
| -    | 流量控制       | 设置迁移任务在指定时段的最大流量带宽。 <ul style="list-style-type: none"> <li>不勾选，则不进行流量控制</li> <li>勾选，进行流量控制。在“开始时间”、“结束时间”和“限速”文本框中输入数值。</li> </ul> <b>须知</b><br>例如：设置开始时间08:00、结束时间12:00、限制最大流量20 MB/s，当迁移任务运行到（08:00~12:00）该时间段时，最大迁移速度将限制为20 MB/s。该时间段以外，迁移速度不做限制。 | <ul style="list-style-type: none"> <li>最多可添加五条流量控制规则。</li> <li>时区时间以控制台所选Region的当地时间为准。</li> </ul> |
|      | 启用LTS服务    | <ul style="list-style-type: none"> <li>启用：记录存储迁移过程中的日志，便于查看和定位迁移中出现的问题。</li> <li>不启用：不记录存储迁移过程中的日志。</li> </ul>                                                                                                                                       | -                                                                                                  |

**步骤5** 参数配置完成后，单击“**确认**”开始创建集群。在迁移集群列表中查看集群状态，状态详情参见**集群状态说明**。如果集群状态为创建失败，将鼠标移动到“创建失败”上，可以查看失败原因。在解决失败原因后，可以通过操作列的“更多 > 重试”功能，重新创建集群。

**步骤6** 迁移集群创建成功后，您可以进行如下操作。

- 当集群状态为**健康**或者**亚健康**时，单击集群操作列的，可以配置存储迁移 workflow。
- 在**集群管理**界面中，您可以添加/删除节点、流量控制规则等。

----结束

## 升级插件版本

当集群的插件版本有更新时，可以一键升级集群插件版本。在集群列表中，单击操作列的“更多 > 升级插件版本”，即可将集群插件升级为最新版本。



## 删除集群

当您完成数据迁移并且集群不再需要时，可以直接删除闲置的集群以释放资源。

**步骤1** 在集群列表，选择需要清理的集群，单击操作列的“更多 > 删除”，弹出**删除集群**窗口。



**步骤2** 选择是否同步删除关联资源后，单击“**确认**”，删除集群。



---结束

## 重试

对于创建失败的集群，在解决失败原因后，可以尝试重新创建集群。在集群列表中，单击操作列的“更多 > 重试”，即可重新创建集群资源。



## 1.9.2 集群管理

在集群创建完成后，您可以根据实际需求随时调整集群内的流量控制规则和节点信息。

### 前提条件

已完成[集群创建](#)。

### 添加流量控制规则

- 步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择创建的[应用迁移项目](#)。
- 步骤2** 在左侧导航栏选择“资源准备>迁移集群”，进入[迁移集群](#)页面。
- 步骤3** 在集群列表中，单击操作列的“管理”，右侧弹出集群详细信息页面。
- 步骤4** 在[流量控制](#)区域，单击“添加”，下方新增一条规则记录。
- 步骤5** 在“开始时间”、“结束时间”和“限速”文本框中输入数值。

例如：设置开始时间08:00、结束时间12:00、限制最大流量20 MB/s，当迁移任务运行到（08:00~12:00）该时间段时，最大迁移速度将限制为20 MB/s。该时间段以外，迁移速度不做限制。

### 须知

- 时区时间以控制台所选Region的当地时间为准。
- 最多可以创建5条记录。

**步骤6** 单击“确认”，完成流量控制规则的添加。

----结束

## 新增节点

新增迁移节点和列举节点的操作步骤相同，以下步骤以新增迁移节点为例进行介绍。

**步骤1** 登录**迁移中心**管理控制台。在左侧导航栏选择“部署>迁移集群”，进入**迁移集群**页面。

**步骤2** 在集群列表中，单击操作列的“管理”，右侧弹出集群详细信息页面。



**步骤3** 在节点信息区域，单击“新增节点>迁移节点”，弹出添加节点页面。



**步骤4** 根据实际需求，选择节点的ECS规格，并输入需要增加的节点数量，节点数量需要同时满足如下要求：

- 迁移节点数+列举节点数+1 ≤100
- 迁移节点数+列举节点数+1 ≤子网可用IP数



**步骤5** 单击“确认”，节点新增完成，系统开始为迁移节点安装迁移插件。添加完成后，可以在节点列表中看到新添加的节点信息。

----结束

### 1.9.3 计费说明

在创建迁移集群时，会产生以下费用：

- 创建Master节点、迁移节点和列举节点会产生ECS服务费用。具体计费详情请参见[ECS按需计费说明](#)或[ECS价格计算器](#)。
- 如果您使用公网迁移时，会产生NAT网关服务费用。计费详情请参见[NAT网关计费说明](#)或[NAT价格计算器](#)。
- 启用LTS服务时，会产生LTS服务费用。具体计费详情请参见[LTS计费说明](#)或[LTS价格计算器](#)。

### 1.9.4 集群状态说明

集群所包含的状态以及状态说明参见[表1-48](#)。

表 1-48 集群状态说明

| 状态   | 说明                                                    |
|------|-------------------------------------------------------|
| 创建中  | 正在创建集群资源                                              |
| 创建失败 | 创建集群资源失败                                              |
| 连接中  | 等待master节点上线，并开始创建和安装其他节点                             |
| 健康   | 所有节点均处于在线状态                                           |
| 亚健康  | 部分迁移节点和列举节点在线，并且至少存在一个迁移节点和一个列举节点在线                   |
| 不可用  | 所有迁移节点或列举节点不在线                                        |
| 离线   | master节点离线，可能是网络中断、ECS删除等原因，请排查Vpcep Service、ECS等资源情况 |
| 升级中  | 升级集群版本中                                               |
| 升级失败 | 升级集群版本失败                                              |
| 废弃中  | 30天内无任务运行，废弃集群中                                       |
| 废弃失败 | 废弃集群失败                                                |
| 删除中  | 删除集群中                                                 |
| 删除失败 | 删除集群失败                                                |
| 已废弃  | 集群已废弃（vpcep删除）                                        |
| 等待创建 | 等待创建集群资源                                              |
| 安装中  | 安装master节点中                                           |
| 安装失败 | 安装master节点失败                                          |
| 等待安装 | 等待安装master节点                                          |
| 等待升级 | 等待升级集群版本                                              |

| 状态   | 说明     |
|------|--------|
| 等待删除 | 等待删除集群 |

## 1.9.5 资源类型和安全组配置说明

本文介绍创建迁移集群所包含的资源类型以及所有节点的安全组配置。

### 迁移集群包含的资源类型和介绍

| 资源类型 | 创建时可配置 | 是否必须 | 规格数量                                                                                                                    | 名称                                                                                                                                                                          | 描述                                         |
|------|--------|------|-------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------|
| ECS  | 是      | 是    | <ul style="list-style-type: none"> <li>Master节点：1台</li> <li>迁移节点：根据创建集群时的参数配置创建</li> <li>列举节点：根据创建集群时的参数配置创建</li> </ul> | “oms_cluster_ecs_” + 节点id                                                                                                                                                   | 承载迁移插件，实现对迁移任务的管理、调度以及数据的迁移。               |
| NAT  | 是      | 否    | 1个                                                                                                                      | “oms_cluster_nat_” + 集群id                                                                                                                                                   | 公网模式下，实现列举节点、迁移节点和源端通信，实现数据的列举和迁移。         |
| LTS  | 是      | 否    | 1个                                                                                                                      | 主机组名称：<br>“oms_lts_host_group_” + 集群id<br>主机组标识：<br>“oms_lts_host_group_label_” + 集群id<br>日志组名称：<br>“oms_lts_log_group_” + 集群id<br>日志流名称：<br>“oms_lts_stream_group_” + 集群id | 保存和搜索迁移集群中的日志。<br>创建集群时未勾选“启用LTS服务”，则不会创建。 |

| 资源类型          | 创建时可配置 | 是否必须 | 规格数量 | 名称                                          | 描述                                                                                              |
|---------------|--------|------|------|---------------------------------------------|-------------------------------------------------------------------------------------------------|
|               |        |      |      | 日志接入名称:<br>“oms_lts_access_confing_” + 集群id |                                                                                                 |
| VPCEP-Service | 否      | 是    | 1个   | regionName + “oms-cluster” + serviceld      | 允许 OMS云服务与迁移集群的 Master节点通信, 实现迁移任务的管理。<br>配置详情请查看 <a href="#">VPCEP-Service 配置</a> 。            |
| 安全组           | 否      | 是    | 2个   | “oms_cluster_sg_master_” + 集群id             | 迁移集群 <b>Master</b> 节点的安全组, 实现与 OMS云服务以及迁移集群内部通信。<br>安全组配置详情请查看 <a href="#">Master 节点安全组配置</a> 。 |

| 资源类型 | 创建时可配置 | 是否必须 | 规格数量 | 名称                             | 描述                                                                           |
|------|--------|------|------|--------------------------------|------------------------------------------------------------------------------|
|      |        |      |      | “oms_cluster_sg_agent_” + 集群id | 迁移集群列举节点和迁移节点的安全组，用于迁移集群内部通信。<br>安全组配置详情请查看 <a href="#">列举节点和迁移节点安全组配置</a> 。 |

## VPCEP-Service 配置

| 配置项    | 详细配置            | 描述                        |
|--------|-----------------|---------------------------|
| 虚拟私有云  | 在创建集群时的网络配置中选择。 | -                         |
| 服务类型   | 接口              | -                         |
| 连接审批   | 是               | 确保终端节点审核后方可使用。            |
| 端口映射   | 22、9002         | 用于集群安装和任务下发、管理。           |
| 后端资源类型 | 云服务器            | 连接Master节点所在的ECS虚拟机。      |
| 连接管理   | OMS云服务的VPCEP端点  | 允许OMS云服务与迁移集群的Master节点通信。 |
| 权限管理   | OMS云服务白名单       | 允许OMS云服务的VPCEP端点进行连接。     |

## Master 节点安全组配置

| 方向  | 优先级 | 策略 | 类型   | 协议端口       | 目的地址/源地址        | 描述                            |
|-----|-----|----|------|------------|-----------------|-------------------------------|
| 入方向 | 1   | 允许 | IPv4 | 全部         | Master节点安全组     | 允许安全组内的虚拟机通信                  |
| 入方向 | 1   | 允许 | IPv6 | 全部         | Master节点安全组     | 允许安全组内的虚拟机通信                  |
| 入方向 | 1   | 允许 | IPv4 | TCP : 22   | 198.19.128.0/17 | 云服OMS云服务和Master节点通信，实现集群安装。   |
| 入方向 | 1   | 允许 | IPv4 | TCP : 9002 | 198.19.128.0/17 | 云服OMS云服务和Master节点通信，实现任务下发、管理 |
| 入方向 | 1   | 允许 | IPv6 | 全部         | 列举和迁移节点安全组      | 允许列举和迁移节点与Master通信            |
| 入方向 | 1   | 允许 | IPv4 | 全部         | 列举和迁移节点安全组      | 允许列举和迁移节点与Master通信            |
| 出方向 | 1   | 允许 | IPv6 | 全部         | ::/128          | 放通出方向流量，实现和源端通信               |
| 出方向 | 1   | 允许 | IPv4 | 全部         | 0.0.0.0/0       | 放通出方向流量，实现和源端通信               |
| 出方向 | 1   | 允许 | IPv6 | 全部         | 列举和迁移节点安全组      | 允许列举和迁移节点与Master通信            |
| 出方向 | 1   | 允许 | IPv4 | 全部         | 列举和迁移节点安全组      | 允许列举和迁移节点与Master通信            |

## 列举节点和迁移节点安全组配置

| 方向  | 优先级 | 策略 | 类型   | 协议端口 | 目的地址/源地址    | 描述                 |
|-----|-----|----|------|------|-------------|--------------------|
| 入方向 | 1   | 允许 | IPv6 | 全部   | Master节点安全组 | 允许列举和迁移节点与Master通信 |

| 方向  | 优先级 | 策略 | 类型   | 协议端口 | 目的地址/源地址    | 描述                 |
|-----|-----|----|------|------|-------------|--------------------|
| 入方向 | 1   | 允许 | IPv4 | 全部   | Master节点安全组 | 允许列举和迁移节点与Master通信 |
| 入方向 | 1   | 允许 | IPv4 | 全部   | 列举和迁移节点安全组  | 允许安全组内的虚拟机通信       |
| 入方向 | 1   | 允许 | IPv6 | 全部   | 列举和迁移节点安全组  | 允许安全组内的虚拟机通信       |
| 出方向 | 1   | 允许 | IPv4 | 全部   | 0.0.0.0/0   | 放通出方向流量，实现和源端通信    |
| 出方向 | 1   | 允许 | IPv6 | 全部   | ::/128      | 放通出方向流量，实现和源端通信    |
| 出方向 | 1   | 允许 | IPv6 | 全部   | Master节点安全组 | 允许列举和迁移节点与Master通信 |
| 出方向 | 1   | 允许 | IPv4 | 全部   | Master节点安全组 | 允许列举和迁移节点与Master通信 |

## 1.10 迁移 workflow

### 1.10.1 workflow 配额限制说明

为确保资源的合理分配和系统的稳定运行，在创建 workflow 时，设定了以下配额限制。请在创建 workflow 前了解这些限制，避免因超出配额而导致创建失败。

- **日创建上限：**在同一项目下，每天允许创建的迁移 workflow 数量上限为50个。
- **运行中 workflow 资源总量限制：**同一用户同时进行迁移的主机数量上限为1000台，超出此数量的主机会在 workflow 第一步暂停并处于等待状态，直至有资源释放或主机完成迁移，等待中的主机将按照创建时间顺序依次自动执行 workflow 后续步骤。
- **单个 workflow 资源限制：**单个 workflow 的资源数量不能超过100个。

### 1.10.2 创建主机迁移 workflow

本节介绍如何使用主机迁移标准模板创建主机迁移 workflow。

#### 注意事项

- 建议单个 workflow 中包含的主机数量不超过100台。
- 同一用户同时进行迁移的主机数量上限为1000台，超出此数量的主机会在 workflow 第一步暂停并处于等待状态，直至有资源释放或主机完成迁移，等待中的主机将按照创建时间顺序依次自动执行 workflow 后续步骤。
- 如果您是首次使用 MgC 的主机迁移 workflow，需要您同意华为云账号进行相关的委托授权。委托权限所包含的细粒度权限请前往[委托权限说明](#)查看。

## 使用须知

在创建主机迁移 workflows 前，请仔细阅读并了解以下使用须知内容。

| 项目        | 须知                                                                                                                                |
|-----------|-----------------------------------------------------------------------------------------------------------------------------------|
| 源端下载带宽    | 用于源端服务器下载SMS-Agent。 <ul style="list-style-type: none"><li>推荐带宽：每台机器不低于30 Mbit/s。</li><li>共用带宽建议：若多台机器共用，平均不低于50 Mbit/s。</li></ul> |
| 迁移带宽      | <ul style="list-style-type: none"><li>重要性：影响数据传输速度和迁移时间。</li><li>推荐值与预估时长：详情请参见<a href="#">迁移需要多长时间?</a></li></ul>                |
| CPU和内存要求  | <ul style="list-style-type: none"><li>内存：至少预留520 MB。</li><li>CPU：Linux系统预留不少于0.3核，Windows系统不低于1核。</li></ul>                       |
| 系统兼容性列表   | 支持迁移的操作系统列表，请参见 <a href="#">兼容性列表</a> 。                                                                                           |
| 主机迁移重要声明  | 了解迁移过程中的重要信息和免责声明，详情请参见 <a href="#">主机迁移服务重要声明有哪些?</a>                                                                            |
| 约束与限制     | 确保迁移过程符合服务条款和限制条件，详情请参见 <a href="#">主机迁移约束与限制</a> 。                                                                               |
| 计费说明      | 了解迁移过程中可能产生的费用，详情请参见 <a href="#">计费说明</a> 。                                                                                       |
| 权限配置      | 使用IAM进行精细的权限管理，详情请参见 <a href="#">权限配置</a> 。                                                                                       |
| 网络配置及迁移端口 | 确保源端和目的端的网络连接和端口开放满足迁移要求，详情请参见 <a href="#">网络配置及端口开放</a> 。                                                                        |

## 前提条件

- 已完成[主机资源采集](#)。
- 已将待迁移的主机资源[关联应用](#)，并对应用内的主机进行[规格评估](#)。规格评估成功后，通过[购买资源](#)，批量购买目的端。如果待迁移应用中的主机已[绑定目的端](#)，则无需进行规格评估。

## 操作步骤

- 步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择创建的[应用迁移项目](#)。
- 步骤2** 在左侧导航栏选择“[迁移实施>迁移工作流](#)”，进入[迁移工作流](#)页面。
- 步骤3** 单击页面右上角的“[创建迁移工作流](#)”按钮，进入[选择模板](#)页面。
- 步骤4** 选择[主机迁移标准模板](#)，单击“[查看步骤详情](#)”，可以查看主机迁移模板所包含的迁移阶段和步骤，以及每个阶段和步骤的详细描述。其中，步骤类型为“[自动](#)”，表示该步由系统自动完成；步骤类型为“[手动](#)”，表示该步需要用户手动操作。单击右下角的“[开始配置](#)”按钮，进行工作流配置。

步骤5 根据表1-49，配置 workflow 参数。

表 1-49 主机迁移 workflow 配置参数说明

| 区域          | 参数     | 说明                                                                                                                                                                                                                                                                                  |
|-------------|--------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| workflow 详情 | 名称     | 用户自定义                                                                                                                                                                                                                                                                               |
|             | 描述     |                                                                                                                                                                                                                                                                                     |
|             | 选择资源范围 | <ul style="list-style-type: none"> <li>按应用：迁移应用内的主机。</li> <li>按迁移方案：对迁移方案中配置的主机进行迁移。</li> </ul>                                                                                                                                                                                     |
| 应用程序        | 应用名称   | 资源范围选择“按应用”时，选择待迁移主机所关联的应用。                                                                                                                                                                                                                                                         |
| 迁移方案        | 方案名称   | 资源范围选择“按迁移方案”时，选择待迁移主机所在的迁移方案。创建迁移方案的方法请参考 <a href="#">配置主机迁移方案</a> 。                                                                                                                                                                                                               |
| 迁移网络配置      | 网络类型   | 选择公网迁移，要求目的端服务器配置有“弹性公网IP（EIP）”，迁移时会使用目的端EIP。                                                                                                                                                                                                                                       |
|             |        | 选择私网迁移，需要提前配置专线、VPN、对等连接或同VPC子网，确保源端能够通过私网访问目的端。 <ul style="list-style-type: none"> <li>如果源端无法访问公网，需要输入源端代理服务器的私网IP和代理软件指定的代理端口。</li> <li>如果源端代理服务器无法访问公网，需要将SMS-Agent安装包放至到源端能访问下载或者借助代理能访问下载的地址。SMS-Agent安装包可以在SMS控制台下载。</li> </ul>                                               |
| 目的端配置       | 区域     | 目的端所在区域，即规格评估时所购买资源的区域。                                                                                                                                                                                                                                                             |
|             | 项目     | 目的端所在区域的项目。                                                                                                                                                                                                                                                                         |
|             | 虚拟私有云  | <ul style="list-style-type: none"> <li>若源端IP是192.168.X.X，则推荐创建的VPC网段是192.168.0.0/16，同时创建一个子网，网段也是192.168.0.0/16。</li> <li>若源端IP是172.16.X.X，则推荐创建的VPC网段是172.16.0.0/12，同时创建一个子网，网段也是172.16.0.0/12。</li> <li>若源端IP是10.X.X.X，则推荐创建的VPC网段是10.0.0.0/8，同时创建一个子网，网段也是10.0.0.0/8。</li> </ul> |
|             | 子网     | 子网网段与虚拟私有云网段相同。                                                                                                                                                                                                                                                                     |

| 区域 | 参数       | 说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|----|----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    | 安全组      | <ul style="list-style-type: none"> <li>Windows系统开放8899端口、8900端口和22端口。</li> <li>Linux系统当前仅支持文件级迁移，需要开放22端口。</li> </ul> <p><b>注意</b></p> <ul style="list-style-type: none"> <li>以上端口，建议只对源端服务器开放。</li> <li>防火墙开放端口与操作系统开放端口保持一致。</li> </ul>                                                                                                                                                                                                                                                                                                                                                                 |
|    | 迁移端口     | <p>默认为22端口。如果22端口被占用，支持将其更改为<b>10000至65535</b>范围内的端口。</p> <p><b>注意</b></p> <p>以下端口为黑名单端口：<br/>135, 139, 445, 5357, 5985, 5986, 8899, 8900, 47001, 49664, 49665, 49666, 49668, 49669, 49671, 49674, 49677, 49679, 49680, 49681, 49682, 49683, 49685, 49690, 49691, 49693, 49694, 49695, 49698, 49699, 49700, 49702, 49703, 49704, 49705, 49706, 49707, 49708, 49709, 49710, 49711, 49712, 49713</p> <p>请确保所填端口未在黑名单中，并且在源端主机上未被其他服务占用，同时网络环境允许该端口的通信。</p> <p>该端口的具体用途如下：</p> <ul style="list-style-type: none"> <li>Linux主机：用于数据传输和控制端口，建立传输通道和文件数据传输。</li> <li>Windows主机：用于传输链路初始化端口，建立数据传输通道。</li> </ul> |
|    | 数据传输控制端口 | Windows主机用于传输任务控制信号的端口，默认值8899，不支持修改。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|    | 数据传输端口   | Windows主机用于传输数据的端口，默认值8900，不支持修改。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|    | 开启加密     | 选择“否”：迁移工作流在创建目的端主机步骤不会设置主机系统盘和数据盘为加密盘。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |

| 区域   | 参数      | 说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|------|---------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|      |         | <p>选择“是”：迁移工作流会在创建目的端主机步骤自动设置主机系统盘和数据盘为加密盘。</p> <p><b>须知</b></p> <ul style="list-style-type: none"> <li>该功能不适用于已绑定目的端的主机。</li> <li>启用该功能后，会对同一工作流内的所有新建目的端主机磁盘进行加密，且采用同一个加密密钥。</li> <li>加密磁盘创建完成后不支持修改。</li> <li>当前密钥共享仅支持授权给账号，不支持授权给用户。</li> </ul> <p>使用该功能，需要先创建委托授权EVS访问KMS。如果您有授权资格，则可直接授权。授权成功后，无需再次授权。如果权限不足，需先联系拥有Security Administrator权限的用户授权，然后再重新操作。授权成功后，设置以下参数：</p> <ul style="list-style-type: none"> <li>从KMS密钥中选择                             <p><b>须知</b></p> <p>只支持AES_256加密算法和启用状态的密钥。</p> <p>在下拉菜单中选择密钥名称。您可以选择使用的密钥如下：</p> <p>默认密钥: 成功授权EVS访问KMS，系统会创建默认密钥“evs/default”。</p> <p>自定义密钥: 即您已有的密钥或者新创建密钥，创建方式请参见<a href="#">创建密钥</a>。</p> </li> <li>输入KMS密钥ID                             <p>输入来自其他用户的共享密钥ID，请确保密钥位于目的端区域。共享方式请参见<a href="#">创建密钥授权</a>。</p> </li> </ul> |
| 高级设置 | 启用一致性校验 | <ul style="list-style-type: none"> <li>选择“是”，在全量复制完成后，系统自动进行数据一致性校验。默认使用快速校验（即比对文件大小和修改时间）。在完成增量同步后，可以再次调整校验策略。</li> <li>选择“否”，在全量复制完成后，不进行数据校验。在完成增量同步后，可以重新设置是否进行一致性校验。</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|      | 迁移后是否开机 | <ul style="list-style-type: none"> <li>选择“否”，迁移完成后目的端服务器自动关机。</li> <li>选择“是”，迁移完成后目的端服务器保持开机状态。</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |

| 区域 | 参数        | 说明                                                                                                                                                                                                                            |
|----|-----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    | 网络限速      | <ul style="list-style-type: none"> <li>选择“否”，代表不限流。</li> <li>选择“是”，根据要迁移的源端带宽大小及业务要求，设置带宽大小。</li> </ul> <p><b>注意</b><br/>当工作流程中迁移的主机数量超过10台，为该工作流分配的带宽资源不足以支持所有迁移任务时，为了避免迁移任务之间的带宽竞争导致部分主机迁移失败，建议您设置一个合理的网络限速，以均衡地分配可用带宽。</p> |
|    | 自动安装rsync | <ul style="list-style-type: none"> <li>选择“否”，不在源端安装rsync组件。</li> <li>选择“是”，在源端自动安装rsync组件。</li> </ul> <p><b>注意</b><br/>Linux迁移依赖rsync组件，若源端未安装rsync组件，可能会导致迁移失败。</p>                                                          |
|    | 保持IP不变    | 主机迁移后保持私有IP不变。使用该功能存在一定的风险，需要您自行评估并承担启用该功能带来的风险。                                                                                                                                                                              |
|    | 启动快速模式    | 适用于无需增量同步的场景，可提高迁移效率。默认关闭，启用后工作流程执行完全量复制步骤后，会直接跳过增量同步步骤，继续执行后续步骤。 <b>请根据您的需求谨慎选择。</b>                                                                                                                                         |
|    | 企业项目      | 选择需要迁移到的企业项目。默认为“default”。                                                                                                                                                                                                    |

**步骤6** 工作流程配置完成后，单击“**下一步：确认并创建**”按钮，进入确认并创建页面。

**步骤7** 确认配置信息无误后，单击“**确认创建工作流**”，弹出**启动工作流程**窗口，此时工作流程已创建完成。

- 如果您希望立即开始迁移，则单击“**确认**”按钮，立即运行工作流。
- 如果您需要在标准工作流程中**添加阶段/步骤**，则单击“**取消**”，此时迁移工作流程状态为“**等待中**”，并未开始迁移。确定工作流程步骤后，单击“**运行**”按钮，开始迁移。

**步骤8** 在迁移工作流程详情页面，可以查看该主机迁移工作流程的**配置信息**和**资源迁移进度**。当资源的迁移进度完成“启动迁移Agent”这一步后，会自动在SMS控制台的迁移服务器列表创建一条迁移任务，关于MgC和SMS主机资源之间的映射关系请查看**MgC主机工作流资源与SMS迁移服务器列表之间的映射关系**。

- 将鼠标移动到主机迁移进度条，在弹窗中可以查看该主机更多迁移信息。
- 当迁移步骤（进度条）到达需要**手动确认**的步骤时，将鼠标移动到进度条，在弹窗中单击**步骤状态**后的“**确认**”，才会继续进行后续迁移步骤。
- 当迁移步骤（进度条）到达“**调整磁盘分区**”步骤时，系统会识别目的端主机是否进行了磁盘扩容。
  - 是，则需要前往**SMS控制台**，对该主机进行“调整磁盘分区”操作，方法请参考**设置目的端**中的**是否调整分区**参数配置说明。调整完成后，单击**步骤状态**后的“**确认**”继续进行后续迁移步骤。

- 否，则跳过该步骤。
- 迁移步骤中的“增量同步”为可重复步骤，在进行业务验证前，可以重复执行。
- 当迁移步骤（进度条）到达“业务割接”时，代表迁移完成。请前往目的端进行业务验证，验证无误后，手动将业务切换到目的端。切换完成后单击“确认”系统将自动进行后续“源端专项清理”和“迁移任务清理”操作。

----结束

## 重试/全量重试

在主机迁移 workflow 运行过程中，MgC Agent（原Edge）会为每台源端服务器启动一个 SMS-Agent 迁移进程，该进程启动后会与 SMS 控制台进行通信，并接收命令执行迁移。在工作流进行到“全量迁移”和“增量同步”两个子步骤阶段，如果源端 SMS-Agent 进程与 SMS 控制台断开连接，MgC 控制台会监控到该情况，同时迁移工作流会失败并出现“重试”和“全量重试”的按钮。以下是导致源端 SMS-Agent 进程与 SMS 控制台断开连接的原因，请根据具体原因选择对应的处理方法：

- **原因一：**因网络异常导致源端 SMS-Agent 进程和 SMS 控制台断开连接。  
该情况通常 SMS-Agent 进程依旧存在，只需要将网络恢复，等待 SMS 控制台的迁移任务恢复连接，然后在迁移工作流单击“重试”即可恢复迁移。

图 1-47 重试



Linux 源端检测 SMS-Agent 进程的指令：

```
ps -ef | grep -v grep | grep linuxmain
```

Windows 源端检测 SMS-Agent 进程的 powershell 指令：

```
Get-Process -Name SMSAgentDeploy -ErrorAction SilentlyContinue
```

- **原因二：**源端主机重启等原因导致 SMS-Agent 进程中止，使得 SMS-Agent 与 SMS 控制台断开连接。

该情况由于 SMS-Agent 进程中止，导致迁移程序无法断点续传，需要已有的 SMS 任务重新创建。针对这种情况，MgC 提供了一键“全量重试”的功能。您只需在迁移工作流单击“全量重试”按钮并确认，工作流将会直接跳到“迁移任务清理”步骤对 SMS 任务进行删除。SMS 任务清理完毕后，工作流会再次从“启动迁移 Agent”步骤运行，并在源端重新启动 SMS-Agent 进程，创建新的 SMS 迁移任务进行迁移。

图 1-48 全量重试



**注意**

全量重试功能会删除旧任务，重新创建SMS迁移任务，已经迁移的数据会被覆盖。

## 增量同步后启用/修改一致性校验策略

当 workflow 执行完“增量同步”步骤后，可以启用/修改一致性校验策略。

**步骤1** 将鼠标移动到进度条，在弹窗中单击一致性校验策略后的“修改”，右侧弹出一致性校验策略窗口。

图 1-49 修改一致性策略



**步骤2** 勾选“数据一致性校验”功能，根据参数说明并结合实际需求配置校验策略。

图 1-50 配置校验策略



表 1-50 校验策略参数说明

| 参数     | 说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|--------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 启用哈希校验 | <p>通过计算和比较文件内容的哈希值进行校验。适合于单文件体积大且文件内容重要的场景，将增加源端CPU负载及源端磁盘IO负载，校验时间也更长。</p> <p><b>注意</b></p> <ul style="list-style-type: none"> <li>被运作中的软件占用的文件无法进行哈希校验。</li> <li>勾选该选项后，必须指定校验目录，并且只校验指定目录下的文件内容。</li> </ul>                                                                                                                                                                                                                                                                                                                                                                     |
| 数据范围   | <ul style="list-style-type: none"> <li>排除下列目录：填写不需要进行数据校验的目录路径，最多支持填写30个路径，路径之间用英文逗号隔开。例如：/root/data,/var。不填则校验所有目录文件。</li> <li>仅包含下列目录：只对填写的目录文件进行校验。</li> </ul> <p><b>须知</b></p> <ul style="list-style-type: none"> <li>请务必输入正确路径，路径错误或路径下不存在文件时，校验结果中会显示校验文件数量为0。</li> <li>需要校验的数据量越大，则一致性校验的用时越长，建议仅对关键目录进行校验。</li> <li>为避免校验数据量过大，以下路径将被默认排除： <ul style="list-style-type: none"> <li>Linux系统的"/bin","/boot","/dev","/home","/etc","/lib","/media","/proc","/sbin","/selinux","/sys","/usr","/var","/run","/tmp"目录。</li> <li>Windows系统的分区根目录，如：C:\，O:\。</li> </ul> </li> </ul> |
| 启用复检模式 | <p>只有进行过至少一次一致性校验后，才可以打开复检选项，启用复检将只对上一次校验结果中不一致的文件进行校验。</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |

**步骤3** 配置完成后，单击“**保存**”按钮，在完成增量同步后，系统自动进行数据一致性校验。

**步骤4** 当同步及校验完成后，前往SMS控制台，[查看校验结果](#)。

----结束

### 1.10.3 创建可用区迁移 workflow

本节介绍如何使用可用区迁移标准模板创建可用区迁移 workflow。

#### 前提条件

- 已完成[主机资源采集](#)。
- 已将待迁移的主机资源[关联应用](#)，并已对应用内的主机进行[规格评估](#)。

#### 操作步骤

- 步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择创建的[应用迁移项目](#)。
- 步骤2** 在左侧导航栏选择“[迁移实施>迁移 workflow](#)”，进入[迁移 workflow](#)页面。
- 步骤3** 单击页面右上角的“[创建迁移 workflow](#)”按钮，进入[选择模板](#)页面。
- 步骤4** 选择[可用区迁移标准模板](#)，单击“[查看步骤详情](#)”，可以查看可用区迁移模板所包含的迁移阶段和步骤，以及每个阶段和步骤的详细描述。其中，步骤类型为“[自动](#)”，表示该步由系统自动完成；步骤类型为“[手动](#)”，表示该步需要用户手动操作。单击“[开始配置](#)”按钮，进行 workflow 配置。
- 步骤5** 根据[表1-51](#)参数说明，配置 workflow 参数。

表 1-51 可用区 workflow 配置参数说明

| 区域          | 参数        | 说明                                                                                                           |
|-------------|-----------|--------------------------------------------------------------------------------------------------------------|
| workflow 详情 | 名称        | 用户自定义                                                                                                        |
|             | 描述        | 用户自定义                                                                                                        |
| 应用程序        | 应用名称      | 选择待迁移资源所关联的应用。                                                                                               |
| 迁移设置        | 区域        | 选择待迁移可用区所在Region，与创建的应用配置保持一致。                                                                               |
|             | 目的可用区     | 选择要迁移到目的端可用区，与创建的应用配置保持一致。                                                                                   |
|             | 网络设置      | 当前只支持“保留源端配置”。                                                                                               |
|             | 目的虚拟机创建策略 | 立即创建<br>workflow 立即开始进行源端备份，源端各自独立备份完成后，立即开始生成镜像和目的端主机。                                                      |
|             | 目的端是否关机   | <ul style="list-style-type: none"> <li>• 选择“是”，生成目的端后，目的端保持关机状态。</li> <li>• 选择“否”，生成目的端后，目的端自动开机。</li> </ul> |

| 区域   | 参数          | 说明                                                                                                                                                         |
|------|-------------|------------------------------------------------------------------------------------------------------------------------------------------------------------|
|      | 增量备份前源端是否关机 | <ul style="list-style-type: none"> <li>选择“是”， workflows在执行增量备份前，会关闭源端主机，这样可以最大限度保证新建目的端数据与源端数据的一致性。</li> <li>选择“否”， workflows在执行增量备份时，依旧保持开机状态。</li> </ul> |
|      | 创建系统盘镜像     | <ul style="list-style-type: none"> <li>选择“是”， 将基于源端系统盘创建系统盘镜像，便于后续目的端重装系统。</li> <li>选择“否”， 不创建系统盘镜像。</li> </ul>                                            |
| 高级设置 | 自动删除镜像      | 选择开启时，业务割接后自动删除迁移过程中产生的备份、快照及镜像文件。                                                                                                                         |
|      | 保持主网卡IP不变   | 选择开启时，将目的端的主网卡私有IP切换为源端主网卡私有IP，并绑定源端主网卡绑定的EIP，此时源端主网卡私有IP将设置为随机IP，并解绑EIP。以上操作不支持自动回退，需要您手动回退。                                                              |

**步骤6** workflows配置完成后，单击“**下一步：确认并创建**”按钮，进入确认并创建页面。

**步骤7** 确认配置信息无误后，单击“**确认创建工作流**”，弹出**启动工作流**窗口，此时工作流已创建完成。

- 如果您希望立即开始迁移，则单击“**确认**”按钮，立即运行工作流。
- 如果您需要在标准工作流中**添加阶段/步骤**，则单击“**取消**”，此时迁移工作流状态为“**等待中**”，并未开始迁移。确定工作流步骤后，单击“**运行**”按钮，开始迁移。

**步骤8** 在迁移工作流详情页面，可以查看该可用区工作流的**配置信息**和**资源迁移进度**。

- 将鼠标移动到主机迁移进度条，在弹窗中可以查看该主机更多迁移信息。
- 当迁移步骤（进度条）到达需要**手动确认**的步骤时，将鼠标移动到进度条，在弹窗中单击**步骤状态**后的“**确认**”，才会继续进行后续迁移步骤。

----结束

## 1.10.4 创建存储迁移工作流

本节介绍如何使用存储迁移标准模板创建存储迁移工作流。

当前支持以下区域（Region）：

- 拉美-圣地亚哥
- 拉美-圣保罗
- 土耳其-伊斯坦布尔
- 亚太-曼谷
- 亚太-新加坡
- 亚太-雅加达
- 中东-利雅得

- 华北-北京四
- 华东-上海一

## 约束与限制

- 关于存储迁移的约束与限制请查看[存储迁移的约束与限制有哪些?](#)
- 源端为阿里云OSS: 当源端对象的存储类型是深度冷归档存储时, 无法使用MgC提供的解冻归档数据功能直接迁移, 需要在源端先手动解冻后再迁移。

## 风险提示

在创建工作流时, 对于源端和目的端的同名对象, 我们提供了三种覆盖策略: **不覆盖**、**直接覆盖**和**条件覆盖**。选择“**不覆盖**”策略进行数据迁移时, 如果迁移任务因错误中断后重启或从暂停状态恢复, 可能会造成部分数据未完全迁移, 而任务显示成功的现象, 影响数据的完整性。为规避这一风险, **请谨慎使用“不覆盖”策略**。

## 前提条件

- 已提前创建[目的端对象桶](#)或[弹性文件系统](#)。
- 已提前[创建迁移集群](#)。

## 操作步骤

- 步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择创建的[应用迁移项目](#)。
- 步骤2** 在左侧导航栏选择“[迁移实施>迁移工作流](#)”, 进入迁移工作流页面。
- 步骤3** 单击页面右上角的“[创建迁移工作流](#)”按钮, 进入[选择模板](#)页面。
- 步骤4** 选择[存储迁移标准模板](#), 单击“[查看步骤详情](#)”, 可以查看存储迁移模板所包含的迁移阶段和步骤, 以及每个阶段和步骤的详细描述。其中, 步骤类型为“**自动**”, 表示该步由系统自动完成。单击“[开始配置](#)”按钮, 进行工作流配置。
- 步骤5** 根据[表1-52](#), 配置基础信息参数。

表 1-52 基础信息参数说明

| 参数 | 说明                                                                                                            |
|----|---------------------------------------------------------------------------------------------------------------|
| 名称 | 用户自定义。                                                                                                        |
| 区域 | 下拉菜单中选择目的端所在区域。                                                                                               |
| 描述 | 用户自定义。                                                                                                        |
| 集群 | 选择用于迁移的集群, 集群中包含master节点、迁移节点和列举节点。如果没有可用集群, 可以 <a href="#">新建集群</a> 。<br><b>须知</b><br>master节点为系统默认创建, 无需配置。 |

- 步骤6** 根据[表1-53](#)、[表1-54](#), 配置源端、目的端信息。

表 1-53 源端信息参数说明

| 参数       | 说明                                                                                                                                                                                                                                                                                                                                                                    | 备注                 |
|----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|
| 数据站点类型   | <p>目前支持迁移的源端存储类型包括：</p> <ul style="list-style-type: none"> <li>● 华为云 OBS</li> <li>● 阿里云 OSS</li> <li>● 百度云 BOS</li> <li>● 腾讯云 COS</li> <li>● 金山云 KS3</li> <li>● 七牛云 KODO</li> <li>● 优刻得 US3</li> <li>● 亚马逊 S3</li> <li>● 微软 Blob</li> <li>● NAS_GFS</li> <li>● NAS_SMB</li> <li>● NAS_NFS_V3_MOUNT</li> <li>● NAS_NFS_V3_PROTOCOL</li> <li>● HTTP/HTTPS数据源</li> </ul> | -                  |
| AK       | 输入源端云账号的访问密钥。                                                                                                                                                                                                                                                                                                                                                         | 数据站点类型选择云平台时填写。    |
| SK       | 输入源端云账号的私有密钥。                                                                                                                                                                                                                                                                                                                                                         |                    |
| 桶        | 输入源端需要迁移的桶名。                                                                                                                                                                                                                                                                                                                                                          |                    |
| 对象存储终端地址 | <p>输入源端桶所在区域的终端节点（Endpoint）信息。</p> <p>例如：站点类型选择阿里云OSS，源端桶所在区域为华东1（杭州），则输入“oss-cn-hangzhou.aliyuncs.com”。</p>                                                                                                                                                                                                                                                          |                    |
| 桶类型      | 根据桶的实际类型选择。桶类型可以通过 <a href="#">查看桶的基本信息</a> 获取。                                                                                                                                                                                                                                                                                                                       | 数据站点类型为华为云OBS时选择。  |
| appId    | <p>填写腾讯云APPID。</p> <p><b>说明</b><br/>您可以在腾讯云控制台账号信息页面查看APPID。</p>                                                                                                                                                                                                                                                                                                      | 数据站点类型为腾讯云COS时填写   |
| 鉴权类型     | <p>鉴权类型的选择支持以下两种：</p> <ul style="list-style-type: none"> <li>● 连接字符串。</li> <li>● 存储账户名称+密钥。</li> </ul> <p>获取以上两种鉴权类型的方式请参考：<a href="#">如何获取微软云的访问凭证</a>。</p>                                                                                                                                                                                                          | 数据站点类型为微软云BLOB时配置。 |

| 参数       | 说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | 备注                                                       |
|----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------|
| 列表路径     | <p>输入存放列表文件的路径。列表文件存放地址必须与目的端桶处于同一区域。将源端待迁移的文件URL和对象名称写入列表文件（每行只能写一个URL和对象名称），列表文件需要提前创建。</p> <p>列表文件约束与限制：</p> <ul style="list-style-type: none"> <li>列表文件类型必须为.txt，其他文件类型不做处理，并且该文件元数据中的“ContentType”只能为：“text/plain”。</li> <li>txt文件行数不超过100000行。</li> <li>单个列表文件大小不能超过300 MB。</li> <li>列表文件存放目录下的列表文件个数不能超过10000个。</li> <li>列表文件必须是UTF-8无BOM格式编码格式。</li> <li>列表文件中每行长度不要超过65535，否则会导致迁移失败。</li> <li>列表文件的元数据中不能设置“ContentEncoding”，否则会导致迁移失败。</li> <li>列表文件中每行使用制表符（键盘上Tab键）\t分隔URL和目的端对象名称，请勿使用空格。格式为：[URL][制表符][目的端对象名称]，其中源端对象名称如果包含中文、特殊字符必须使用URL Encode对URL编码；目的端对象名称如果包含中文、特殊字符也需要使用URL Encode编码。</li> <li>列表文件中每行不要添加无效空格，否则会将空格作为对象名，导致迁移失败。</li> </ul> <p><b>说明</b><br/>出于安全考虑，若您的源端同时支持HTTP和HTTPS协议，建议迁移时使用HTTPS协议；若使用HTTP协议，会存在数据可能会被监听窃取、数据被篡改、中间人重放攻击等风险。</p> | 数据站点类型选择HTTP/HTTPS数据源时填写。                                |
| 文件存储服务地址 | 输入源端文件系统的挂载地址，格式为：IP地址+:/或IP地址+:/ xxx，例如：192.1.1.1:/或192.1.1.1:/0001                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 数据站点类型选择NAS_SMB/NAS_NFS_V3_MOUNT/NAS_NFS_V3_PROTOCOL时填写。 |
| 路径       | 输入待迁移文件路径，格式为：/文件夹。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                                          |

| 参数         | 说明                                                             | 备注                           |
|------------|----------------------------------------------------------------|------------------------------|
| 用户名        | 输入具有访问源端文件系统所有文件权限的用户名，例如：administrator。                       | 数据站点类型选择 <b>NAS_SMB</b> 时填写。 |
| 密码         | 输入用户名对应的密码。                                                    |                              |
| Windows上的域 | 输入对应的域名。<br><b>说明</b><br>只需输入.com之前的内容，例如：域名为test.com，则输入test。 |                              |

表 1-54 目的端信息参数说明

| 参数       | 说明                                                                                                                                            | 备注                                            |
|----------|-----------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------|
| 数据站点类型   | 根据源端存储类型，选择华为云对应的存储类型。                                                                                                                        | -                                             |
| AK       | 输入目的端华为云账号的访问密钥。                                                                                                                              | 数据站点类型选择 <b>华为云OBS</b> 时填写。                   |
| SK       | 输入目的端华为云账号的私有密钥。                                                                                                                              |                                               |
| 桶        | 选择目的端OBS桶。                                                                                                                                    |                                               |
| 对象存储终端地址 | 输入目的端OBS桶所在区域的终端节点（Endpoint）信息。例如：目的端桶所在区域为 <b>华北-北京四</b> ，则输入“obs.cn-north-4.myhuaweicloud.com”。<br><b>说明</b><br>可在OBS桶的基本信息查看。              |                                               |
| 指定前缀     | 将对象迁移到指定前缀，例如：源端对象路径为/A/B/C.txt，目的端前缀配置为/D，迁移后目的端对象路径为/D/A/B/C.txt。详细介绍请参考 <a href="#">设置“保存到指定前缀”参数</a> 中的功能介绍。                              |                                               |
| 文件存储服务地址 | 输入目的端文件系统的挂载地址。挂载地址可以在文件系统列表页面，单击 <b>挂载地址</b> 后面的  图标复制获取。 | 数据站点类型选择 <b>NAS_SMB/NAS_NFS_V3_MOUNT</b> 时填写。 |
| 路径       | 输入目的端接收文件的路径，格式为：/文件夹。                                                                                                                        |                                               |
| 用户名      | 输入具有访问目的端文件系统所有文件权限的用户名，例如：administrator。                                                                                                     | 数据站点类型选择 <b>NAS_SMB</b> 时填写。                  |
| 密码       | 输入用户名对应的密码。                                                                                                                                   |                                               |

| 参数         | 说明                                                             | 备注 |
|------------|----------------------------------------------------------------|----|
| Windows上的域 | 输入对应的域名。<br><b>说明</b><br>只需输入.com之前的内容，例如：域名为test.com，则输入test。 |    |

**步骤7** 根据表1-55，进行迁移配置。

**表 1-55** 迁移配置参数说明

| 参数     | 子参数  | 说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|--------|------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 迁移任务类型 | 全量迁移 | 迁移源端桶/文件夹内的全部数据。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|        | 列表迁移 | 通过列表文件过滤迁移，只迁移列表文件中的对象。<br>在“ <b>列表路径</b> ”输入存放在目的端桶的对象列表文件路径。<br>对象列表文件的约束与限制如下： <ul style="list-style-type: none"> <li>● 目的端为OBS                             <ul style="list-style-type: none"> <li>- 对象列表文件大小不能超过30 MB。</li> <li>- 对象列表文件类型必须为.txt类型，并且元数据ContentType只能为text/plain。</li> <li>- 对象列表文件必须是UTF-8无BOM格式编码格式。</li> <li>- 对象列表文件每行只能包含一个对象名称，且对象名称使用URL Encode编码。</li> <li>- 对象列表文件行长度不超过16KB，否则迁移失败。</li> <li>- 对象列表文件元数据中不能设置ContentEncoding，否则迁移失败。</li> <li>- 对象列表文件行数不超过10000行。</li> </ul> </li> <li>● 目的端为NAS                             <ul style="list-style-type: none"> <li>- 对象列表文件大小不能超过30 MB。</li> <li>- 对象列表文件类型必须为.txt类型。</li> <li>- 对象列表文件必须是UTF-8无BOM格式编码格式。</li> <li>- 对象列表文件每行只能包含一个对象名称，且对象名称使用URL Encode编码。</li> <li>- 对象列表文件行长度不超过16KB，否则迁移失败。</li> <li>- 对象列表文件行数不超过10000行。</li> </ul> </li> </ul> |

| 参数     | 子参数           | 说明                                                                                                                                                                                                                                                                                           |
|--------|---------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|        | 前缀迁移          | <p>仅支持云平台数据类型。</p> <p>在“前缀路径”输入对象名称或者名称前缀，则只迁移与输入对象名称或名称前缀一致的对象。</p> <p><b>须知</b></p> <ul style="list-style-type: none"> <li>输入对象前缀路径时，如果对象文件在源端桶的根目录下，则直接添加前缀；如果对象文件存在目录路径，则需要在对象前缀前添加对象文件所在的目录路径。格式为：文件夹/对象前缀</li> <li>多个前缀之间用英文逗号分隔。</li> </ul>                                           |
| 并发子任务数 | -             | <p>用户自定义。配置的数目不能超过可用迁移节点数x10。例如：可用的迁移节点数为2，则最大子任务数配置数目需要≤20。</p>                                                                                                                                                                                                                             |
| 同名覆盖方式 | 不覆盖           | <p>对于源端和目的端存在的同名文件，跳过源端文件，保留目的端文件。</p> <p><b>警告</b></p> <ul style="list-style-type: none"> <li>首次迁移时选择“不覆盖”，不会迁移源端中父文件夹的属性到目的端，可能会造成文件夹属性不完整。为规避这一风险，首次迁移时请谨慎使用“不覆盖”策略。</li> <li>选择“不覆盖”策略进行数据迁移时，如果迁移任务因错误中断后重启或从暂停状态恢复，可能会造成部分数据未完全迁移，而任务显示成功的现象，影响数据的完整性。为规避这一风险，请谨慎使用“不覆盖”策略。</li> </ul> |
|        | 直接覆盖          | <p>对于源端和目的端存在的同名文件，源端文件直接覆盖目的端文件。</p>                                                                                                                                                                                                                                                        |
|        | 对象大小与最后修改时间覆盖 | <ul style="list-style-type: none"> <li>对于文件迁移：源端和目的端存在的同名文件，保留最后修改时间较新的文件；如果最后修改时间相同，比较文件大小，大小不同则源端覆盖目的端文件；如果最后修改时间和大小都相同，则跳过；</li> <li>对于文件夹迁移：当文件迁移完成后再比对文件夹大写、时间和权限，如果文件夹大小不同/时间不相同/权限不相同，则源端覆盖目的端文件夹；</li> </ul> <p><b>说明</b><br/>对于空文件夹，覆盖策略与文件相同。</p>                                  |
|        | CRC 64对比覆盖    | <ul style="list-style-type: none"> <li>迁移前，对于源端和目的端存在的同名对象，对比源端和目的端对象元数据中的CRC64值，相同则跳过，不同则覆盖。如果源端与目的端对象元数据中不存在CRC64值，则默认使用大小/最后修改时间对比覆盖策略执行。</li> </ul> <p><b>说明</b></p> <ul style="list-style-type: none"> <li>目前仅支持华为云、阿里云、腾讯云。</li> <li>目的端OBS桶需开启CRC64特性白名单。</li> </ul>                  |

| 参数    | 子参数           | 说明                                                                                                                                                                                             |
|-------|---------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 一致性校验 | 对象大小与最后修改时间校验 | 默认配置，通过对比源端和目的端对象大小与最后修改时间，判断迁移后数据是否完整。                                                                                                                                                        |
|       | CRC 64校验      | 通过对比源端和目的端对象元数据中CRC64值，判断迁移后数据是否完整。如果源端与目的端对象元数据中不存在CRC64值，则默认使用大小/最后修改时间校验策略。<br><b>说明</b> <ul style="list-style-type: none"> <li>目前仅支持华为云、阿里云、腾讯云。</li> <li>目的端OBS桶需开启CRC64特性白名单。</li> </ul> |
| 迁移元数据 | -             | 选择是否迁移元数据。 <ul style="list-style-type: none"> <li>勾选，则迁移源端对象元数据。</li> <li>不勾选，仅迁移Content-Type、Content-Encoding元数据。</li> </ul>                                                                  |
| 清理集群  | -             | 选择工作流是否创建清理集群步骤。 <ul style="list-style-type: none"> <li>勾选，迁移工作流将自动创建清理集群步骤，您还可以继续配置是否删除NAT网关、安全组、VPCEP服务等。</li> <li>不勾选，迁移工作流不创建清理集群步骤。</li> </ul>                                            |

步骤8 （可选）根据表1-56，配置高级选项。

表 1-56 高级选项参数说明

| 参数      | 说明                                                                                                                                                                                                                                        | 备注                          |
|---------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------|
| 目的端存储类型 | 选择迁移到华为云OBS后，对象的存储类型。存储策略说明请参见 <a href="#">存储类型介绍</a> 。                                                                                                                                                                                   | -                           |
| KMS加密   | <ul style="list-style-type: none"> <li>不勾选，迁移后对象状态与目的端桶级服务端加密状态保持一致。</li> <li>勾选，所有对象在迁移到OBS桶后都将被加密存储。</li> </ul> <b>说明</b> <ul style="list-style-type: none"> <li>使用KMS服务端加密可能会降低迁移速率10%左右。</li> <li>目的端桶所在区域支持KMS加密时，此参数才可用。</li> </ul> | 仅目的端选择 <b>华为云OBS</b> 时支持配置。 |

| 参数        | 说明                                                                                                                                                                                                                                                                                                                                                                                           | 备注 |
|-----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 解冻归档数据    | <ul style="list-style-type: none"> <li>不勾选：如果遇到归档类型对象，则直接判定该对象迁移失败并跳过，继续迁移其余对象。</li> <li>勾选：如果遇到归档类型对象，则自动解冻该对象并进行迁移。如果解冻失败，则判定该对象迁移失败并跳过，继续迁移其余对象。</li> </ul> <p><b>说明</b><br/>在数据迁移时，系统会对所有数据发起取回操作，源端云服务商收取相应的API调用费用和存储费用。</p>                                                                                                                                                           | -  |
| 过滤配置      | 通过黑、白名单以及时间段过滤迁移文件。过滤策略详细介绍请参见 <a href="#">过滤策略说明</a> 。                                                                                                                                                                                                                                                                                                                                      |    |
| CDN迁移支持   | <p>当默认域名不能满足您的迁移需求，比如您需要一个自定义域名以便使用CDN降低源端出口下载流量的费用，可以勾选此选项。在域名输入框中输入自定义域名，并选择传输协议。HTTP协议安全级别低于HTTPS协议，推荐使用HTTPS协议。</p> <p>源端为阿里云OSS、腾讯云COS时，如果使用了CDN鉴权，还需要选择“鉴权类型”并输入“鉴权密钥”。</p>                                                                                                                                                                                                            |    |
| SMN消息发送配置 | <p>选择是否使用SMN消息通知功能，该功能可以通知您迁移任务的结果。</p> <ul style="list-style-type: none"> <li>不勾选，不接收SMN消息。</li> <li>勾选，接收SMN消息。选择SMN主题，会向该主题的订阅者发送消息；选择消息发送语言；选择发送消息的触发条件，支持同时勾选。</li> </ul>                                                                                                                                                                                                               |    |
| 流量控制      | <p>设置迁移任务在指定时段的最大流量带宽。</p> <ul style="list-style-type: none"> <li>不勾选，不进行流量控制。</li> <li>勾选，进行流量控制。分别设置“开始时间”、“结束时间”和“限速”数值。<br/>例如：设置开始时间08:00、结束时间12:00、限速为20 MB/s，当迁移任务运行到（08:00~12:00）该时间段时，最大迁移速度将限制为20 MB/s。该时间段以外，迁移速度不做限制。</li> </ul> <p><b>说明</b></p> <ul style="list-style-type: none"> <li>限速的数值范围：0~1048576 MB/s。</li> <li>最多可以创建5条记录。</li> <li>时区时间以目的端Region的当地时间为准。</li> </ul> | -  |

| 参数   | 说明                                                                                                                                                                                                                                                                                                                                                        | 备注 |
|------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 定时迁移 | <p>设置迁移任务在指定时间内自动迁移。</p> <ul style="list-style-type: none"> <li>不勾选，手动启动/暂停迁移任务。</li> <li>勾选，设置任务在指定时间范围内进行迁移，指定时间以外则暂停迁移。<br/>例如： <ul style="list-style-type: none"> <li>设置开始时间08:00、结束时间12:00，则迁移任务会在每天早晨08:00至中午12:00该时间段进行迁移。该时间段以外，则迁移任务暂停。</li> <li>设置开始时间12:00、结束时间08:00，则迁移任务会在每天中午12:00至次日早晨08:00该时间段进行迁移。该时间段以外，则迁移任务暂停。</li> </ul> </li> </ul> | -  |

**步骤9** workflows配置完成后，单击“**下一步：确认并创建**”按钮，进入确认并创建页面。

**步骤10** 确认配置信息无误后，单击“**确认创建工作流**”，弹出**启动工作流**窗口，此时工作流已创建完成。

- 如果您希望立即开始迁移，则单击“**确认**”按钮，立即运行工作流。
- 如果您需要在标准工作流中**添加阶段/步骤**，则单击“**取消**”，此时迁移工作流状态为“**等待中**”，并未开始迁移。确定工作流步骤后，单击“**运行**”按钮，开始迁移。

**步骤11** 在工作流详情页面，可以查看该存储迁移工作流的**配置信息**和**资源迁移进度**。还可以进行如下操作：

- 将鼠标移动到资源迁移进度条，在弹窗中可以查看该资源更多迁移信息。
- 当迁移步骤（进度条）到达需要**手动确认**的步骤时，您需要将鼠标移动到进度条上，并在弹出的窗口中单击**步骤状态**后的“**确认**”。只有完成这一操作，系统才会继续进行后续的迁移步骤。
- 在**基础信息**区域，单击集群名称后的“**管理**”，右侧弹出集群详细信息页面。在这个页面上，您可以进行如下操作：
  - 设置流量控制规则**：您可以添加、编辑或删除流量控制规则，以确保集群内的流量按照您的需求进行管理。
  - 节点管理**：您可以根据实际需求，添加/删除迁移节点和列举节点，也可以对现有节点进行插件升级。

**步骤12** （可选）单击资源迁移进度条或进度条弹窗中的“**资源迁移进度详情**”，右侧弹出该资源迁移详情页面，可以查看**任务概览**和**进度详情**等信息。还可以进行如下操作：

| 操作                 | 步骤                                                                                                                                                                                                                                                                                                                                                                               |
|--------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 更换集群               | <p>只有OMS任务状态（即 workflow 状态）为“已暂停”时，可以进行“更换集群”操作。</p>  <p>1. 在<b>任务概览</b>区域，单击迁移集群名称后的“<b>更换</b>”。</p> <p>2. 在下拉列表中选择新集群，单击“<b>确认</b>”，更换集群后，系统开始自动运行。</p>                                                                                                                                      |
| 修改定时迁移             | <p>1. 在<b>任务概览</b>区域，单击<b>定时迁移</b>后的“<b>修改</b>”。</p> <p>2. 设置<b>开始时间</b>和<b>结束时间</b>，单击“<b>确认</b>”。</p>                                                                                                                                                                                                                                                                          |
| 修改并发子任务数           | <p>1. 在<b>进度详情</b>区域，单击期望并发子任务数下方的“<b>修改</b>”，配置期望并发的子任务数。配置要求：配置的数目不能超过可用迁移节点数x10。例如，可用的迁移节点数为2，则配置数目需要≤20。</p> <p>2. 单击“<b>确认</b>”，完成并发子任务数修改。</p>                                                                                                                                                                                                                             |
| 添加迁移限速规则           | <p>1. 在<b>迁移速度</b>区域，单击“<b>添加</b>”按钮，可以增加一条限速规则，用于限制该资源在某个时间段内的迁移速度。</p> <p><b>须知</b></p> <ul style="list-style-type: none"> <li>- 限速取值范围：1 MB~1024 GB。</li> <li>- 创建多条限速规则时，限速时段不能重叠。<br/>例如：已有一条规则，限速时段是8:00~12:00，那么7:00~13:00，7:00~8:00，9:00~12:00...有时间重叠的规则，都无法创建成功。</li> <li>- 开始时间不能晚于等于结束时间。<br/>例如：不能设置时间段为23:00~01:00。</li> </ul> <p>2. 设置完成后，单击“<b>保存</b>”，完成添加。</p> |
| 获取迁移失败/跳过/成功资源列表文件 | <p>在<b>文件统计</b>区域，可以查看存放迁移<b>失败/跳过/成功</b>的资源列表文件路径。单击路径，可以进入存放文件的OBS桶，获取对应列表文件。</p>                                                                                                                                                                                                                                                                                              |
| 查看迁移流量统计           | <p>在<b>流程统计</b>区域，可以查看该资源近1小时、近6小时、近24小时、整个周期迁移流量使用情况。</p>                                                                                                                                                                                                                                                                                                                       |

----结束

## 过滤策略说明

过滤策略参数填写通配符规则说明和约束说明参见[表1-57](#)。

表 1-57 过滤策略参数说明

| 参数  | 说明                                                                                                            | 填写规则                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | 约束说明                                                                                                                                                                                                                                                                |
|-----|---------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 黑名单 | 命中黑名单规则，则文件不进行迁移/一致性对比。支持精确匹配和模糊匹配。                                                                           | <ul style="list-style-type: none"> <li>精确匹配<br/>全路径精确匹配，填写文件绝对路径，特殊字符用反斜杠 (\) 转义。</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | <ul style="list-style-type: none"> <li>除{和}以外，不能出现连续的规则符号，例如：<br/>***、*?、**?、?*、?*、*[*、*]*、*}?、?[*、*]、{*}、{*}、{*、*}、*、*。</li> <li>{和}之间的模式，只能使用*做通配符。</li> <li>{和}之间的模式，不能有{}嵌套。</li> <li>文件绝对路径在黑名单和白名单同时命中，则黑名单生效，即不做迁移。</li> <li>需要配置多个匹配规则时，用英文分号隔开。</li> </ul> |
| 白名单 | <ul style="list-style-type: none"> <li>白名单为空时，迁移所有文件。</li> <li>白名单不为空时，仅对绝对路径命中白名单规则的文件进行迁移/一致性对比。</li> </ul> | <p><b>注意</b><br/>配置黑、白名单注意事项：</p> <ul style="list-style-type: none"> <li>如果文件存储服务地址是以冒号加斜杠 (:/) 结尾，请确保在黑、白名单中填写的路径是相对于挂载地址的绝对路径。<br/>例如，文件存储服务地址为 "192.1.1.1:/", 挂载地址为 "/mnt/turbo", 则填写 "/mnt/turbo" 下的绝对路径。<br/>挂载信息示例：<br/>[root@oms-cluster-ecs filter_test]# pwd<br/>/mnt/sts_turbo/mgc/filter_test<br/>[root@oms-cluster-ecs filter_test]# ll<br/>drwxr-xr-x 2 root root 0 Aug 16 15:27 test2<br/>-rw-r--r-- 1 root root 5 Aug 16 15:27 test2.log<br/>如果要使黑名单精确匹配以避免迁移 "test2.log" 文件，则应在黑名单中填写路径： /mgc/filter_test/test2.log</li> <li>如果文件存储服务地址在冒号加斜杠 (:/) 之后还包含其他路径，配置黑、白名单时，除了填写挂载地址下的绝对路径外，还要加上文件存储服务地址:/之后的部分。<br/>例如，文件存储服务地址为 "192.1.1.1:/mgc-test", 挂载地址为 "/mnt/turbo", 则填写 "mgc-test+/mnt/turbo" 下的绝对路径。<br/>挂载信息示例：<br/>[root@oms-cluster-ecs execution-service]# cd /mnt/turbo/autotest/filter_test/<br/>[root@oms-cluster-ecs filter_test]# ll<br/>-rw-r--r-- 1 root root 14 Aug 8 09:22 test1.log<br/>drwxr-xr-x 1 root root 4096 Aug 8 09:22 test2<br/>-rw-r--r-- 1 root root 14 Aug 8 09:22 test2.log<br/>如果要使黑名单精确匹配以避免迁移 "test2" 文件，则应在黑名单中填写路径： /mgc-test/autotest/filter_test/test2</li> </ul> |                                                                                                                                                                                                                                                                     |
|     |                                                                                                               | <ul style="list-style-type: none"> <li>模糊匹配</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                                                                                                                                                                                                                                                     |

| 参数  | 说明                                                                         | 填写规则                                                                                                                                                                                                                                                                 | 约束说明 |
|-----|----------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
|     |                                                                            | <ul style="list-style-type: none"> <li>- *匹配任何字符0个或者多个，但不匹配斜杠(/)</li> <li>- 使用**匹配任何字符0个或者多个，包括斜杠(/)</li> <li>- ?匹配除斜杠(/)之外的任何字符，只能匹配一个</li> <li>- {和}定义元素之间的选择：包含一个以逗号分隔的模式列表，其中任何一种模式都可能匹配，可以包含通配符。</li> <li>- 特殊字符可以在它们之前使用\进行转义，\后面为通配符时，表示转义，否则表示\本身</li> </ul> |      |
| 时间段 | 配置时间段，根据配置，迁移最后修改时间在配置时间段内的文件或者文件夹。<br>开始时间和结束时间可以配置为空，为空时表示不限制，时间可以精确到分钟。 |                                                                                                                                                                                                                                                                      |      |

过滤策略场景规则示例以及举例见下表。

### 须知

例如第一级目录为“test”，那么：

- NAS的路径格式为：“/test”。
- OBS的路径格式为：“test”。

| 场景              | NAS规则示例                                  | OBS规则示例                                | 说明             |
|-----------------|------------------------------------------|----------------------------------------|----------------|
| 文件路径以xx结尾(特定后缀) | /xx /**xx                                | xx/**xx                                | xx可以是包含*和?的表达式 |
| 根路径下第一层xx开始的文件  | /xx*<br>• /testssss 匹配<br>• /test/xx 不匹配 | /xx*<br>• testssss 匹配<br>• test/xx 不匹配 |                |
| 文件路径以xx开始       | /xx**<br>/xx/**                          | xx**<br>xx/**                          |                |
| 文件名包含xx         | **xx*                                    | **xx*                                  |                |
| 文件路径包含xx        | **xx**                                   | **xx**                                 |                |

| 场景                  | NAS规则示例     | OBS规则示例     | 说明                 |
|---------------------|-------------|-------------|--------------------|
| 文件路径以xx开始，以yy结尾     | /xx**yy     | xx**yy      | xx, yy可以是包含*和?的表达式 |
| 文件路径以xx, yy结尾(特定后缀) | **{xx,yy}   | **{xx,yy}   |                    |
| 文件名包含xx, yy         | **{xx,yy}*  | **{xx,yy}*  |                    |
| 文件路径包含xx, yy        | **{xx,yy}** | **{xx,yy}** |                    |

## 1.10.5 创建对象存储批量迁移 workflow

本节介绍如何使用对象存储批量迁移模板，批量创建对象存储数据迁移任务，实现对对象存储数据的高效迁移。

### 约束与限制

- 关于存储迁移的约束与限制请查看[存储迁移的约束与限制有哪些?](#)
- 源端为阿里云OSS：当源端对象的存储类型是深度冷归档存储时，无法使用MgC提供的解冻归档数据功能直接迁移，需要在源端先手动解冻后再迁移。

### 风险提示

在创建工作流时，对于源端和目的端的同名对象，我们提供了三种覆盖策略：**不覆盖**、**直接覆盖**和**条件覆盖**。选择“**不覆盖**”策略进行数据迁移时，如果迁移任务因错误中断后重启或从暂停状态恢复，可能会造成部分数据未完全迁移，而任务显示成功的现象，影响数据的完整性。为规避这一风险，**请谨慎使用“不覆盖”策略**。

### 前提条件

已完成[对象存储迁移方案配置](#)。

### 操作步骤

- 步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择创建的[应用迁移项目](#)。
- 步骤2** 在左侧导航栏选择“[迁移实施>迁移工作流](#)”，进入[迁移工作流](#)页面。
- 步骤3** 单击页面右上角的“[创建迁移工作流](#)”按钮，进入[选择模板](#)页面。
- 步骤4** 选择[对象存储批量迁移模板](#)，单击“[查看步骤详情](#)”，可以查看该模板所包含的迁移阶段和步骤，以及每个阶段和步骤的详细描述。其中，步骤类型为“**自动**”，表示该步由系统自动完成。单击“[开始配置](#)”按钮，进行[工作流配置](#)。
- 步骤5** 在[基础信息](#)区域，输入自定义的工作流**名称**和**描述**。
- 步骤6** 在[迁移方案](#)区域，选择[创建的对象存储迁移方案](#)。选择方案后，可以看到该方案的配置概览。单击“[查看详情](#)”，可以查看该方案的更多信息。



**步骤7** 在**迁移集群**区域，选择迁移使用的集群。默认使用迁移方案中配置的集群，您也可以在下拉列表中重新选择集群，选择后仅对当前工作流生效，不会修改迁移方案中配置的集群。

**步骤8** 根据表1-58，进行迁移配置。

表 1-58 迁移配置参数说明

| 参数     | 子参数           | 说明                                                                                                                                                                                                                                                                                |
|--------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 并发子任务数 | -             | 用户自定义。配置的数目不能超过可用迁移节点数x10。例如：可用的迁移节点数为2，则最大子任务数配置数目需要≤20。                                                                                                                                                                                                                         |
| 同名覆盖方式 | 不覆盖           | 对于源端和目的端存在的同名文件，跳过源端文件，保留目的端文件。<br><b>警告</b> <ul style="list-style-type: none"> <li>首次迁移时选择“不覆盖”，不会迁移源端中父文件夹的属性到目的端，可能会造成文件夹属性不完整。为规避这一风险，首次迁移时请谨慎使用“不覆盖”策略。</li> <li>选择“不覆盖”策略进行数据迁移时，如果迁移任务因错误中断后重启或从暂停状态恢复，可能会造成部分数据未完全迁移，而任务显示成功的现象，影响数据的完整性。为规避这一风险，请谨慎使用“不覆盖”策略。</li> </ul> |
|        | 直接覆盖          | 对于源端和目的端存在的同名文件，源端文件直接覆盖目的端文件。                                                                                                                                                                                                                                                    |
|        | 对象大小与最后修改时间覆盖 | <ul style="list-style-type: none"> <li>对于文件迁移：源端和目的端存在的同名文件，保留最后修改时间较新的文件；如果最后修改时间相同，比较文件大小，大小不同则源端覆盖目的端文件；如果最后修改时间和大小都相同，则跳过；</li> <li>对于文件夹迁移：当文件迁移完成后再比对文件夹大写、时间和权限，如果文件夹大小不同/时间不相同/权限不相同，则源端覆盖目的端文件夹；</li> </ul> <b>说明</b><br>对于空文件夹，覆盖策略与文件相同。                               |
|        | CRC64对比覆盖     | <ul style="list-style-type: none"> <li>迁移前，对于源端和目的端存在的同名对象，对比源端和目的端对象元数据中的CRC64值，相同则跳过，不同则覆盖。如果源端与目的端对象元数据中不存在CRC64值，则默认使用大小/最后修改时间对比覆盖策略执行。</li> </ul> <b>说明</b> <ul style="list-style-type: none"> <li>目前仅支持华为云、阿里云、腾讯云。</li> <li>目的端OBS桶需开启CRC64特性白名单。</li> </ul>              |

| 参数    | 子参数           | 说明                                                                                                                                                                                             |
|-------|---------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 一致性校验 | 对象大小与最后修改时间校验 | 默认配置，通过对比源端和目的端对象大小与最后修改时间，判断迁移后数据是否完整。                                                                                                                                                        |
|       | CRC64 校验      | 通过对比源端和目的端对象元数据中CRC64值，判断迁移后数据是否完整。如果源端或目的端对象元数据中不存在CRC64值，则默认使用大小/最后修改时间校验策略。<br><b>说明</b> <ul style="list-style-type: none"> <li>目前仅支持华为云、阿里云、腾讯云。</li> <li>目的端OBS桶需开启CRC64特性白名单。</li> </ul> |
| 迁移元数据 | -             | 选择是否迁移元数据。 <ul style="list-style-type: none"> <li>勾选，则迁移源端对象元数据。</li> <li>不勾选，仅迁移Content-Type、Content-Encoding元数据。</li> </ul>                                                                  |

**步骤9** （可选）根据[表1-59](#)，配置高级选项。

**表 1-59** 高级选项参数说明

| 参数      | 说明                                                                                                                                                                                                                                        |
|---------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 报告级别    | 勾选需要的报告级别，迁移完成后自动生成对应级别的报告，并上传目的端/oms目录下。支持多选。<br>如：勾选成功对象，迁移完成后自动生成成功级别的报告，并上传目的端/oms目录下。                                                                                                                                                |
| 是否增量迁移  | 选择“否”，不增量迁移。                                                                                                                                                                                                                              |
|         | 选择“是”，配置增量迁移同名覆盖方式和增量迁移模式，详细配置方法参考 <a href="#">配置增量迁移</a> 。                                                                                                                                                                               |
| 目的端存储类型 | 选择迁移到华为云OBS后，对象的存储类型。存储策略说明请参见 <a href="#">存储类型介绍</a> 。                                                                                                                                                                                   |
| KMS加密   | <ul style="list-style-type: none"> <li>不勾选，迁移后对象状态与目的端桶级服务端加密状态保持一致。</li> <li>勾选，所有对象在迁移到OBS桶后都将被加密存储。</li> </ul> <b>说明</b> <ul style="list-style-type: none"> <li>使用KMS服务端加密可能会降低迁移速率10%左右。</li> <li>目的端桶所在区域支持KMS加密时，此参数才可用。</li> </ul> |

| 参数        | 说明                                                                                                                                                                                                                                                                                                                                                                                           |
|-----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 解冻归档数据    | <ul style="list-style-type: none"> <li>不勾选：如果遇到归档类型对象，则直接判定该对象迁移失败并跳过，继续迁移其余对象。</li> <li>勾选：如果遇到归档类型对象，则自动解冻该对象并进行迁移。如果解冻失败，则判定该对象迁移失败并跳过，继续迁移其余对象。</li> </ul> <p><b>说明</b><br/>在数据迁移时，系统会对所有数据发起取回操作，源端云服务商会收取相应的API调用费用和存储费用。</p>                                                                                                                                                          |
| 过滤配置      | 通过黑、白名单以及时间段过滤迁移文件。过滤策略详细介绍请参见 <a href="#">过滤策略说明</a> 。                                                                                                                                                                                                                                                                                                                                      |
| CDN迁移支持   | <p>当默认域名不能满足您的迁移需求，比如您需要一个自定义域名以便使用CDN降低源端出口下载流量的费用，可以勾选此选项。在域名输入框中输入自定义域名，并选择传输协议。HTTP协议安全级别低于HTTPS协议，推荐使用HTTPS协议。</p> <p>源端为阿里云OSS、腾讯云COS时，如果使用了CDN鉴权，还需要选择“鉴权类型”并输入“鉴权密钥”。</p>                                                                                                                                                                                                            |
| SMN消息发送配置 | <p>选择是否使用SMN消息通知功能，该功能可以通知您迁移任务的结果。</p> <ul style="list-style-type: none"> <li>不勾选，不接收SMN消息。</li> <li>勾选，接收SMN消息。选择SMN主题，会向该主题的订阅者发送消息；选择消息发送语言；选择发送消息的触发条件，支持同时勾选。</li> </ul>                                                                                                                                                                                                               |
| 流量控制      | <p>设置迁移任务在指定时段的最大流量带宽。</p> <ul style="list-style-type: none"> <li>不勾选，不进行流量控制。</li> <li>勾选，进行流量控制。分别设置“开始时间”、“结束时间”和“限速”数值。<br/>例如：设置开始时间08:00、结束时间12:00、限速为20 MB/s，当迁移任务运行到（08:00~12:00）该时间段时，最大迁移速度将限制为20 MB/s。该时间段以外，迁移速度不做限制。</li> </ul> <p><b>说明</b></p> <ul style="list-style-type: none"> <li>限速的数值范围：0~1048576 MB/s。</li> <li>最多可以创建5条记录。</li> <li>时区时间以目的端Region的当地时间为准。</li> </ul> |

**步骤10** workflows配置完成后，单击“**下一步：确认并创建**”按钮，进入确认并创建页面。

**步骤11** 确认配置信息无误后，单击“**确认创建工作流**”，弹出**启动工作流**窗口，此时工作流已创建完成。

- 如果您希望立即开始迁移，则单击“**确认**”按钮，立即运行工作流。
- 如果您需要在 workflows 中**添加阶段/步骤**，则单击“**取消**”，此时迁移工作流状态为“**等待中**”，并未开始迁移。确定 workflows 步骤后，单击“**运行**”按钮，开始迁移。

**步骤12** 在 workflow 详情页面，可以查看该 workflow 的**配置信息**和**资源迁移进度**。还可以进行如下操作：

- 将鼠标移动到资源迁移进度条，在弹窗中可以查看该资源更多迁移信息。
- 当迁移步骤（进度条）到达需要**手动确认**的步骤时，您需要将鼠标移动到进度条上，并在弹出的窗口中单击**步骤状态**后的“**确认**”。只有完成这一操作，系统才会继续进行后续的迁移步骤。
- 在**基础信息**区域，单击集群名称后的“**管理**”，右侧弹出集群详细信息页面。在这个页面上，您可以进行如下操作：
  - **设置流量控制规则**：您可以添加、编辑或删除流量控制规则，以确保集群内的流量按照您的需求进行管理。
  - **节点管理**：您可以根据实际需求，添加/删除迁移节点和列举节点，也可以对现有节点进行插件升级。
- 在**基础信息**区域，单击展开“**高级选项**”，当开启增量迁移并且选择自动迁移模式时，自动增量迁移次数支持修改。

----结束

## 配置增量迁移

增量迁移配置说明参见下表。

**表 1-60** 增量迁移配置

| 参数         | 说明                                                                                                                                                                                                                                                                                                    |
|------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 增量迁移同名覆盖方式 | 与 <b>迁移配置</b> 中的同名覆盖方式相同。<br><b>说明</b><br>仅在增量迁移阶段生效。                                                                                                                                                                                                                                                 |
| 增量迁移模式     | <ul style="list-style-type: none"> <li>• <b>手动</b>：手动触发增量迁移。首次增量迁移会自动进行，完成后可以再次手动触发。</li> <li>• <b>自动</b>：根据设置的增量迁移周期，增量迁移次数上限和时区触发增量迁移。<br/>举例：选择自动，如下图配置，定时任务则会在每周周一按（GMT+8）时区的早上8点开始执行增量迁移，一共执行2次。</li> </ul>  |

- 选择**手动**，当对象存储 workflow 执行到**批量创建增量迁移任务**时，会立即执行增量迁移。
- 选择**自动**，当对象存储 workflow 执行到**批量创建增量迁移任务**时，您可以进行如下操作：
  - **手动增量迁移**：在等待自动增量迁移的过程中，可以手动立即执行增量迁移。

- 跳过增量迁移：直接跳过增量迁移，执行业务验证。

## 1.10.6 创建文件存储批量迁移 workflow

本节介绍如何使用文件存储批量迁移模板，批量创建文件存储数据迁移任务，实现文件存储数据的高效迁移。

### 约束与限制

关于文件存储迁移的约束与限制请查看[存储迁移的约束与限制有哪些？](#)

### 风险提示

在创建工作流时，对于源端和目的端的同名对象，我们提供了三种覆盖策略：**不覆盖**、**直接覆盖**和**条件覆盖**。选择“**不覆盖**”策略进行数据迁移时，如果迁移任务因错误中断后重启或从暂停状态恢复，可能会造成部分数据未完全迁移，而任务显示成功的现象，影响数据的完整性。为规避这一风险，**请谨慎使用“不覆盖”策略**。

### 前提条件

已完成[文件存储批量迁移方案配置](#)。

### 操作步骤

- 步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择创建的[应用迁移项目](#)。
- 步骤2** 在左侧导航栏选择“[迁移实施>迁移工作流](#)”，进入迁移工作流页面。
- 步骤3** 单击页面右上角的“[创建迁移工作流](#)”按钮，进入[选择模板](#)页面。
- 步骤4** 选择[文件存储批量迁移模板](#)，单击“[查看步骤详情](#)”，可以查看该模板所包含的迁移阶段和步骤，以及每个阶段和步骤的详细描述。其中，步骤类型为“**自动**”，表示该步由系统自动完成。单击“[开始配置](#)”按钮，进行工作流配置。



- 步骤5** 在[基础信息](#)区域，输入自定义的工作流名称和描述。
- 步骤6** 在[迁移方案](#)区域，选择[创建的文件存储迁移方案](#)。选择方案后，可以看到该方案的配置概览。单击“[查看详情](#)”，可以查看该方案的更多信息。



**步骤7** 在**迁移集群**区域，选择迁移使用的集群。默认使用迁移方案中配置的集群，您也可以在下拉列表中重新选择集群，选择后仅对当前工作流生效，不会修改迁移方案中配置的集群。

**步骤8** 根据**表1-61**，进行迁移配置。

**表 1-61** 迁移配置参数说明

| 参数     | 子参数  | 说明                                                                                                                                                                                                                                                                                |
|--------|------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 并发子任务数 | -    | 用户自定义。配置的数目不能超过可用迁移节点数x10。例如：可用的迁移节点数为2，则最大子任务数配置数目需要≤20。                                                                                                                                                                                                                         |
| 同名覆盖方式 | 不覆盖  | 对于源端和目的端存在的同名文件，跳过源端文件，保留目的端文件。<br><b>警告</b> <ul style="list-style-type: none"> <li>首次迁移时选择“不覆盖”，不会迁移源端中父文件夹的属性到目的端，可能会造成文件夹属性不完整。为规避这一风险，首次迁移时请谨慎使用“不覆盖”策略。</li> <li>选择“不覆盖”策略进行数据迁移时，如果迁移任务因错误中断后重启或从暂停状态恢复，可能会造成部分数据未完全迁移，而任务显示成功的现象，影响数据的完整性。为规避这一风险，请谨慎使用“不覆盖”策略。</li> </ul> |
|        | 直接覆盖 | 对于源端和目的端存在的同名文件，源端文件直接覆盖目的端文件。                                                                                                                                                                                                                                                    |
|        | 条件覆盖 | <ul style="list-style-type: none"> <li>对于文件迁移：源端和目的端存在的同名文件，保留最后修改时间较新的文件；如果最后修改时间相同，比较文件大小，大小不同则源端覆盖目的端文件；如果最后修改时间和大小都相同，则跳过；</li> <li>对于文件夹迁移：当文件迁移完成后再比对文件夹大写、时间和权限，如果文件夹大小不同/时间不相同/权限不相同，则源端覆盖目的端文件夹；</li> </ul> <b>说明</b><br>对于空文件夹，覆盖策略与文件相同。                               |
| 迁移元数据  | -    | 选择是否迁移元数据。 <ul style="list-style-type: none"> <li>勾选，则迁移源端对象元数据。</li> <li>不勾选，仅迁移Content-Type、Content-Encoding元数据。</li> </ul>                                                                                                                                                     |

**步骤9** （可选）根据**表1-62**，配置高级选项。

表 1-62 高级选项参数说明

| 参数        | 说明                                                                                                                                                                                                                                                                                                                                                                                           |
|-----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 割接前是否增量同步 | <ul style="list-style-type: none"> <li>选择是，会在迁移阶段的业务验证步骤后增加自动执行的“增量迁移”步骤。</li> <li>选择否，在迁移阶段的业务验证步骤执行完后，直接进入业务割接阶段。</li> </ul>                                                                                                                                                                                                                                                               |
| 过滤配置      | 通过黑、白名单以及时间段过滤迁移文件。过滤策略详细介绍请参见 <a href="#">过滤策略说明</a> 。                                                                                                                                                                                                                                                                                                                                      |
| SMN消息发送配置 | <p>选择是否使用SMN消息通知功能，该功能可以通知您迁移任务的结果。</p> <ul style="list-style-type: none"> <li>不勾选，不接收SMN消息。</li> <li>勾选，接收SMN消息。选择SMN主题，会向该主题的订阅者发送消息；选择消息发送语言；选择发送消息的触发条件，支持同时勾选。</li> </ul>                                                                                                                                                                                                               |
| 流量控制      | <p>设置迁移任务在指定时段的最大流量带宽。</p> <ul style="list-style-type: none"> <li>不勾选，不进行流量控制。</li> <li>勾选，进行流量控制。分别设置“开始时间”、“结束时间”和“限速”数值。<br/>例如：设置开始时间08:00、结束时间12:00、限速为20 MB/s，当迁移任务运行到（08:00~12:00）该时间段时，最大迁移速度将限制为20 MB/s。该时间段以外，迁移速度不做限制。</li> </ul> <p><b>说明</b></p> <ul style="list-style-type: none"> <li>限速的数值范围：0~1048576 MB/s。</li> <li>最多可以创建5条记录。</li> <li>时区时间以目的端Region的当地时间为准。</li> </ul> |

**步骤10** workflows配置完成后，单击“**下一步：确认并创建**”按钮，进入确认并创建页面。

**步骤11** 确认配置信息无误后，单击“**确认创建工作流**”，弹出**启动工作流**窗口，此时工作流已创建完成。

- 如果您希望立即开始迁移，则单击“**确认**”按钮，立即运行工作流。
- 如果您需要在 workflows 中**添加阶段/步骤**，则单击“**取消**”，此时迁移工作流状态为“**等待中**”，并未开始迁移。确定 workflows 步骤后，在 workflows 列表单击操作列“**运行**”按钮，开始迁移。

**步骤12** 单击 workflows 名称，进入 workflows 详情页面，可以查看该 workflows 的**配置信息**和**资源迁移进度**。还可以进行如下操作：

- 将鼠标移动到资源迁移进度条，在弹窗中可以查看该资源更多迁移信息。
- 当迁移步骤（进度条）到达需要**手动确认**的步骤时，您需要将鼠标移动到进度条上，并在弹出的窗口中单击**步骤状态**后的“**确认**”。只有完成这一操作，系统才会继续进行后续的迁移步骤。
- 在**基础信息**区域，单击集群名称后的“**管理**”，右侧弹出集群详细信息页面。在这个页面上，您可以进行如下操作：

- **设置流量控制规则：**您可以添加、编辑或删除流量控制规则，以确保集群内的流量按照您的需求进行管理。
- **节点管理：**您可以根据实际需求，添加/删除迁移节点和列举节点，也可以对现有节点进行插件升级。

----结束

## 1.10.7 添加阶段/步骤

您可以在工作流标准模板的迁移阶段/步骤，添加卡点，对迁移阶段/步骤进行控制。

### ⚠ 注意

- 只有迁移阶段/步骤状态为**等待中/暂停**，才可以选择该阶段/步骤进行添加。迁移阶段/步骤处于**运行中/完成**状态时，无法添加。
- 如果上一阶段/步骤处于**运行中/暂停/完成**状态，该阶段/步骤只能选择**添加后置阶段/步骤**。
- **可重复**类型的步骤，无法选择**添加后置步骤**。

### 添加阶段

- 步骤1** 在**迁移工作流详情**页面，将鼠标移动到迁移阶段名称上，在弹出的窗口中，选择**添加前置阶段/添加后置阶段**，弹出**添加阶段**窗口。
- 步骤2** 为迁移阶段自定义**名称和说明**，单击“**添加步骤**”，选择步骤类型，为步骤自定义名称和说明，单击“**确认**”按钮。支持添加多个步骤。
- 步骤3** 迁移阶段中的步骤添加完成后，单击“**确认**”按钮，完成迁移阶段的添加。

### 须知

只有自行添加的阶段，可以进行修改、删除操作，模板中的迁移阶段无法进行修改和删除。

----结束

### 添加步骤

- 步骤1** 在**迁移工作流详情**页面，将鼠标移动到步骤名称上，选择**添加前置步骤/添加后置步骤**，弹出**添加步骤**窗口。
- 步骤2** 根据**表1-63**，选择步骤类型，为步骤自定义名称和说明，单击“**确认**”按钮，添加步骤完成。

表 1-63 步骤类型说明

| 步骤类型 | 说明              |
|------|-----------------|
| 卡点   | 迁移到该步骤时，需要手动确认。 |

**步骤3** 返回迁移阶段，查看添加的步骤。

### 须知

只有自行添加的步骤，可以进行修改、删除操作，模板中的步骤无法进行修改和删除。

----结束

## 1.11 大数据数据迁移（MaxCompute 迁移到 DLI）

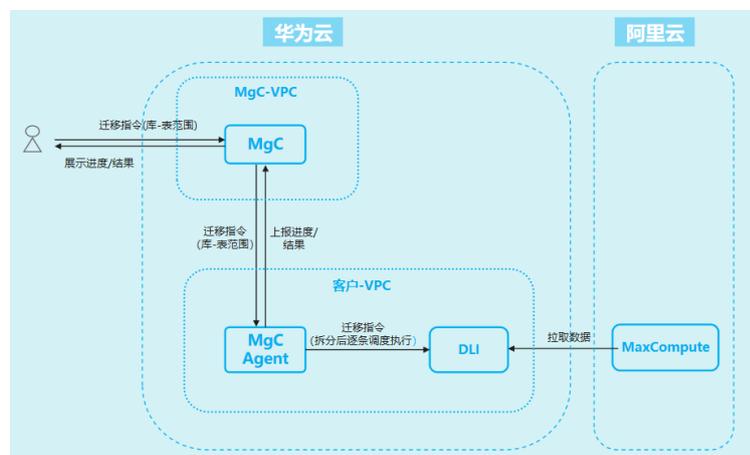
### 1.11.1 大数据迁移概述

#### 迁移场景

将阿里云 MaxCompute 数据迁移至华为云数据湖探索（DLI）。

#### 数据迁移原理图

图 1-51 大数据数据迁移原理图



## 迁移操作流程

图 1-52 大数据数据迁移流程图



### 1.11.2 使用须知

#### 约束与限制

- MaxCompute不支持datetime类型数据格式，在迁移之前，需要将所有datetime类型的字段转换为字符串（STRING）格式。
- MaxCompute不支持varchar和char类型字段，在迁移之前，需要将所有varchar和char类型的字段转换为字符串（STRING）格式。
- 元数据迁移不支持事务表（即DDL中带有transactional=true字段的表）。
- 由于阿里云最近更新了timestamp\_ntz字段类型，目前该字段类型尚未被阿里云SDK所兼容支持，这一兼容性问题导致SDK无法识别并读取该字段，从而使该字段所在的表元数据迁移失败。
- 在数据迁移时，如果重复迁移分区数据时提示错误信息："Do not support update partition SerDelInfo for DLI table"，需要您手动删除DLI中的分区数据，然后再进行迁移。
- 在进行**指定分区数据迁移**的场景下，由于MaxCompute在创建表时会自动将表名转换为小写，因此在迁移过程中，请确保Excel文件中的库表名称与MaxCompute中的实际表名一致（全部小写）。如果Excel文件中填写的库表名称为大写，迁移任务会因为无法正确识别表名导致无法迁移。

#### 注意事项

- **CSV模板文件编辑**  
当使用MgC提供的CSV模板文件进行数据筛选时，请不要使用Excel对CSV模板文件进行编辑，通过Excel编辑并保存的模板文件无法被MgC识别。
- **DLI队列连接选择**  
**元数据迁移**和**数据迁移**需要使用不同的DLI队列连接，**元数据迁移**需要使用**DLI SQL队列连接**，**数据迁移**需要使用**DLI 通用队列连接**。配置错误的队列连接类型

会导致迁移任务失败，因此在配置任务时，必须仔细检查并确保任务类型与队列类型相匹配。

- **元数据迁移与库的数量**

在进行元数据迁移时，MgC会根据识别到的源端数据库的数量，在DLI中自动创建相应的库。如果是整库迁移，通常只会创建一个库；而表迁移可能需要创建多个库。通常，DLI的库数量上限为10。如果DLI空间不足，可能会导致元数据迁移失败，因此在元数据迁移前，需要评估源端库的数量，确保DLI有足够的空间来创建新的库。

### 1.11.3 准备工作

为了确保迁移过程的顺利进行，需要完成以下必要的准备工作。

#### 华为账号准备

使用MgC之前，您需要拥有一个华为账号或一个可用于访问MgC的IAM用户，并获取账号/IAM用户的访问密钥（AK/SK）。获取方法请参考[准备工作](#)。

##### 说明

在创建元数据连接时，如果要使用“采集资源使用信息”功能和查看资源概览，需要为IAM账号的AK/SK添加MRS服务和DLI服务的只读权限（ReadOnly）。

#### 阿里云访问密钥（AK/SK）准备

准备一组源端阿里云账号的Access Key（AK）和Secret Key（SK），获取方法请参考[查看RAM用户的AccessKey信息](#)。

确保该AK/SK具备以下最小权限要求：

- 拥有OSS服务的只读权限：**AliyunReadOnlyAccess**
- 拥有MaxCompute服务的只读权限：**AliyunMaxComputeReadOnlyAccess**

配置权限的方法参考[为RAM用户授权](#)。

（可选）如果迁移的元数据存在分区表，需要为迁移账号开通Information Schema权限，方法请参考[RAM用户访问授权](#)。

#### 创建大数据迁移项目

在MgC控制台为本次迁移创建独立的项目，方法请参考[项目管理](#)。

#### 配置委托权限与服务授权

为了确保正常使用DLI的功能，需要配置一条包含DLI权限和OBS权限的委托权限。

**步骤1** 登录华为云管理控制台。

**步骤2** 单击右上方登录的用户名，在下拉列表中选择“统一身份认证”。

**步骤3** 在左侧导航栏中，单击“委托”。

**步骤4** 在“委托”页面，单击“创建委托”。



**步骤5** 在“创建委托”页面，设置如下参数：

- 委托名称：用户自定义，例如“dli\_obs\_agency\_access”。
- 委托类型：选择“云服务”。
- 云服务：在下拉列表中选择“数据湖探索 DLI”。
- 持续时间：按实际需求选择。
- 描述：非必选。



**步骤6** 配置完委托的基本信息后，单击“下一步”，进入选择策略页签。

**步骤7** 单击右上角的“新建策略”，参考**步骤8**和**步骤9**，分别创建一条OBS策略和一条DLI策略。如果已创建的授权策略中有所需的权限策略，则无需创建，直接选择已有策略即可。

**步骤8** 配置策略信息。

1. 策略名称：用户自定义，例如：dli-obs-agency。
2. 策略配置方式：选择“JSON视图”。
3. 在策略内容中粘贴以下自定义策略。

**注意**

请替换“bucketName”为存放jar包的桶名称。

```
{
 "Version": "1.1",
 "Statement": [
 {
 "Effect": "Allow",
 "Action": [
 "obs:bucket:GetBucketPolicy",
```

```

"obs:bucket:GetLifecycleConfiguration",
"obs:bucket:GetBucketLocation",
"obs:bucket:ListBucketMultipartUploads",
"obs:bucket:GetBucketLogging",
"obs:object:GetObjectVersion",
"obs:bucket:GetBucketStorage",
"obs:bucket:GetBucketVersioning",
"obs:object:GetObject",
"obs:object:GetObjectVersionAcl",
"obs:object:DeleteObject",
"obs:object:ListMultipartUploadParts",
"obs:bucket:HeadBucket",
"obs:bucket:GetBucketAcl",
"obs:bucket:GetBucketStoragePolicy",
"obs:object:AbortMultipartUpload",
"obs:object:DeleteObjectVersion",
"obs:object:GetObjectAcl",
"obs:bucket:ListBucketVersions",
"obs:bucket:ListBucket",
"obs:object:PutObject"
],
"Resource": [
"OBS:*:*:bucket:bucketName",//请替换bucketName为存放jar包的桶名称
"OBS:*:*:object:*"
]
},
{
"Effect": "Allow",
"Action": [
"obs:bucket:ListAllMyBuckets"
]
}
]
}

```

### 步骤9 配置DLI策略信息。

1. 策略名称：用户自定义，例如：dli-agency。
2. 策略配置方式：选择“JSON视图”。
3. 在策略内容中粘贴以下自定义策略。

```

{
"Version": "1.1",
"Statement": [
{
"Effect": "Allow",
"Action": [
"dli:table:showPartitions",
"dli:table:alterTableAddPartition",
"dli:table:alterTableAddColumns",
"dli:table:alterTableRenamePartition",
"dli:table:delete",
"dli:column:select",
"dli:database:dropFunction",
"dli:table:insertOverwriteTable",
"dli:table:describeTable",
"dli:database:explain",
"dli:table:insertIntoTable",
"dli:database:createDatabase",
"dli:table:alterView",
"dli:table:showCreateTable",
"dli:table:alterTableRename",
"dli:table:compaction",
"dli:database:displayAllDatabases",
"dli:database:dropDatabase",
"dli:table:truncateTable",
"dli:table:select",
"dli:table:alterTableDropColumns",
"dli:table:alterTableSetProperties",

```

```
"dli:database:displayAllTables",
"dli:database:createFunction",
"dli:table:alterTableChangeColumn",
"dli:database:describeFunction",
"dli:table:showSegments",
"dli:database:createView",
"dli:database:createTable",
"dli:table:showTableProperties",
"dli:database:showFunctions",
"dli:database:displayDatabase",
"dli:table:alterTableRecoverPartition",
"dli:table:dropTable",
"dli:table:update",
"dli:table:alterTableDropPartition"
]
 }
] }
}
```

**步骤10** OBS和DLI自定义策略创建完成后，单击“**下一步**”，返回委托授权页面。

**步骤11** 勾选新建的OBS和DLI自定义策略，单击“**下一步**”，授权范围方案选择“**所有资源**”。

**步骤12** 单击“**确定**”，完成授权。授权后需等待15-30分钟才可生效。

**步骤13** 参考[更新委托权限](#)，将DLI委托权限进行更新。

----结束

## 创建虚拟私有云 VPC

购买ECS时，需要进行网络配置，可提前创建VPC和子网，方法请参考[创建虚拟私有云和子网](#)。

### 注意

ECS的VPC网段不能和DLI队列（弹性资源池）使用的网段冲突，DLI队列默认使用的网段为172.16.0.0/18。

## 购买 ECS

- **购买ECS**：在与目的端DLI同一区域（Region）下购买一台Linux系统的ECS主机，购买方法请参考[购买弹性云服务器](#)。其中网络配置选择上一步[创建的虚拟私有云和子网](#)。该ECS主机还需要满足以下条件：
  - 可以访问外网（公网），并检查是否能够访问MgC和IoTDA服务的域名，具体需要检查的域名请查看[域名列表](#)。
  - 安全组的出方向规则中，需要包含8883端口。
  - 操作系统为：**CentOS 8.x**。
  - 推荐规格不小于**8U16G**。
- **创建并绑定EIP**：ECS主机要实现访问外网，需要配置EIP（弹性公网IP）。如果主机已绑定EIP则跳过该操作无需重复绑定；如果未绑定EIP请参考[申请EIP](#)和[将EIP绑定至ECS](#)进行绑定。
  - 计费模式：建议选择**按需计费**。
  - 带宽大小：推荐**5 Mbit/s**。

## 安装 MgC Agent（原 Edge）并连接迁移中心

- **安装MgC Agent**：在购买的ECS主机上，安装用于配合数据校验的MgC Agent，方法请参见[安装Linux版本的MgC Agent](#)。
- **注册MgC Agent用户**：在浏览器中访问Linux主机的网卡IP地址+端口（27080），如：<https://x.x.x.x:27080>。首次访问后会进入注册页面，分别输入用户名、密码，确认密码无误后，单击《隐私政策声明》，仔细阅读后，勾选“我已阅读并同意”，最后单击“确认”即可完成注册。
- **连接迁移中心**：安装并注册成功后，将MgC Agent与迁移中心进行连接，连接方法请参考[连接迁移中心](#)。

### 注意

- MgC Agent与迁移中心连接时，请将输入的华为云账号AK/SK保存为目的端凭证。
- MgC Agent不具备自动重启功能，严禁在任务执行期间重启MgC Agent，以避免任务失败。

## 添加凭证

在MgC Agent上分别添加一条MaxCompute凭证（源端阿里云账号的AK/SK）和数据湖DLI凭证（目的端华为云账号的AK/SK），认证方式选择AK/SK，添加方法请参考[添加资源凭证](#)。

- MaxCompute凭证输入源端阿里云账号的AK/SK。

### 创建凭证

|        |                                         |
|--------|-----------------------------------------|
| * 使用类型 | <input type="text" value="大数据"/>        |
| * 类型   | <input type="text" value="MaxCompute"/> |
| * 凭证名称 | <input type="text"/>                    |
| * 认证方式 | <input type="text" value="AK/SK"/>      |
| * ak   | <input type="text" value="请输入ak"/>      |
| * sk   | <input type="text" value="请输入sk"/>      |

取消

确认

- 数据湖DLI凭证输入目的端华为云账号的AK/SK。

### 创建凭证

|        |                                          |
|--------|------------------------------------------|
| * 使用类型 | <input type="text" value="大数据"/>         |
| * 类型   | <input type="text" value="数据湖探索 (DLI)"/> |
| * 凭证名称 | <input type="text"/>                     |
| * 认证方式 | <input type="text" value="AK/SK"/>       |
| * ak   | <input type="text" value="请输入ak"/>       |
| * sk   | <input type="text" value="请输入sk"/>       |

## 创建 OBS 桶并上传 jar 包

在华为云OBS创建一个存储桶，并将数据迁移过程中所依赖的 Java 文件（jar包）上传至该OBS桶。创建OBS桶的方法请参考[创建桶](#)。上传文件的方法请参考[上传对象](#)。

数据迁移所依赖的jar包分别为：migration-dli-spark-1.0.0.jar、fastjson-1.2.54.jar、datasource.jar。这三个jar包是数据迁移过程中不可或缺的，三个jar包的用途和获取方法如下：

- **migration-dli-spark-1.0.0.jar**
  - 用途：用于创建spark会话并提交sql语句。
  - 获取方法：在MgC Agent所部署主机的“/opt/cloud/Edge/tools/plugins/collectors/bigdata-migration/dliSpark”路径下获取。
- **fastjson-1.2.54.jar**：
  - 用途：用于处理JSON格式的数据交换。
  - 获取方法：在MgC Agent所部署主机的“/opt/cloud/Edge/tools/plugins/collectors/bigdata-migration/deltaSpark”路径下获取。
- **datasource.jar**：
  - 用途：包含数据源的配置和连接逻辑，允许服务连接到不同的数据库或数据存储系统。
  - 获取方法：需要您根据需求自行获取并编译。

## 购买弹性资源池并添加队列

弹性资源池为DLI作业运行提供所需的计算资源（CPU和内存）。购买方法请参考[创建弹性资源池](#)。

创建完弹性资源池后，需要为弹性资源池分别添加一条通用队列和一条SQL队列，用于后续作业的运行。添加方法请参考[添加队列](#)。

## 创建 DLI 增强型跨源连接

**步骤1** 登录DLI管理控制台，在左侧导航栏单击“资源管理 > 弹性资源池”，可进入弹性资源池管理页面。

**步骤2** 单击弹性资源池名称前的 ，在展开的信息中，获取该弹性资源池的网段。

**步骤3** 将弹性资源池的网段加入ECS安全组的入方向规则中。

1. 登录弹性云服务器控制台。
2. 在弹性云服务器列表中，单击购买的Linux主机名称（用于安装MgC Agent的主机）。
3. 选择**安全组**页签，单击安全组的“配置规则”。



4. 选择**入方向规则**，单击“添加规则”。



5. 设置入方向规则参数。
  - 优先级：设置为1。
  - 策略：选择“允许”。
  - 类型：选择“IPv4”。
  - 协议端口：选择“全部协议”。
  - 源地址：选择IP地址，并填写弹性资源池的网段。
6. 单击“确定”完成安全组配置。

**步骤4** 在DLI管理控制台的左侧导航栏单击“跨源管理”，进入跨源管理页面。

**步骤5** 选择“增强型跨源”，单击“创建”，弹出创建连接窗口。



**步骤6** 根据**表1-64**，配置增强型跨源连接信息。

表 1-64 参数配置说明

| 参数    | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
|-------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 连接名称  | 用户自定义                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| 弹性资源池 | 选择 <b>购买的弹性资源池</b> 。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| 虚拟私有云 | 选择 <b>创建的虚拟私有云和子网</b> 。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| 子网    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| 路由表   | 保持默认。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| 主机信息  | <p>需要添加两条源端MaxCompute侧主机信息，填写内容和格式如下：</p> <p><b>EndpointIP Endpoint</b>（中间有空格）</p> <p><b>TunnelEndpointIP TunnelEndpoint</b>（中间有空格）</p> <p>两条主机信息之间用回车进行隔开。举例：</p> <p>118.178.xxx.xx service.cn-<br/>hangzhou.maxcompute.aliyun.com.vipgds.alibabadns.com</p> <p>47.97.xxx.xx dt.cn-hangzhou.maxcompute.aliyun.com</p> <p><b>EndpointIP和TunnelEndpointIP的获取方法：</b>使用任一有公网IP的主机，分别对源端MaxCompute所在地域的Endpoint和Tunnel Endpoint执行ping指令，在回显中即可获得对应IP地址。</p> <p>Endpoint和Tunnel Endpoint信息请查看<b>各地域Endpoint对照表</b>。</p> |

**步骤7** 单击“确定”，创建增强型跨源连接。创建完成后，增强型跨源连接的连接状态显示“已激活”，代表该连接创建成功。

---结束

## 添加并配置路由

- **添加路由**

为**创建的DLI增强型跨源连接**添加**两条**路由，方法请参考**添加路由**。路由IP与创建跨源连接时，配置的主机信息IP保持一致。

- **配置路由**

1. 登录虚拟私有云管理控制台。
2. 在左侧导航栏，选择“虚拟私有云 > 路由表”，进入路由表列表页面。
3. 在路由表列表中，找到创建跨源连接时对应的路由表（即ECS所在VPC的路由表），单击路由表名称，进入路由表详情页面。
4. 单击“添加路由”，弹出添加路由窗口。



5. 按照提示配置参数。需要添加两条路由，单击 **继续添加**，可以依次增加多条路由。

- 目的地址类型：选择“IP地址”。
- 目的地址：与创建跨源连接时，配置的主机信息IP保持一致。
- 下一跳类型：选择“服务器实例”。
- 下一跳：选择**购买的ECS**。



## 配置 SNAT 转换规则

配置完SNAT转换规则后，如果ECS重启，则需要重新配置。

**步骤1** 登录**购买的ECS主机**。

**步骤2** 依次执行下列两条命令：

```
sysctl net.ipv4.ip_forward=1
```

该命令用于启用 Linux 主机上的 IP 转发功能。

```
iptables -t nat -A POSTROUTING -o eth0 -s {DLI弹性资源池所在网段} -j SNAT --to {ECS私网IP}
```

设置iptables规则以进行网络地址转换。

----**结束**

## 测试 DLI 队列与数据源的连通性

- 步骤1** 登录DLI管理控制台，在左侧导航栏单击“资源管理 > 队列管理”，可进入队列管理页面。
- 步骤2** 找到为弹性资源池添加的DLI队列，单击操作列的“更多 > 测试地址连通性”，弹出测试地址连通性窗口。
- 步骤3** 参考窗口中的样例，依次输入MaxCompute侧的“Endpoint域名”、“Endpoint IP”、“Tunnel Endpoint域名”和“Tunnel Endpoint IP”进行四次连通性测试，端口建议输入443端口。



- 步骤4** 单击“测试”。

如果测试地址可连通，页面上将提示地址可达。

如果测试地址不可连通，页面上将提示地址不可达，请检查网络配置后重试。检查网络配置即检查所测试的VPC对等连接或跨源连接是否处于已激活状态。

---结束

## 开通 DLI 服务 Spark 3.3 和 jar 程序访问元数据白名单

数据迁移需要使用DLI服务的**Spark 3.3**特性和**jar程序访问DLI元数据**特性，请联系DLI服务的技术支持人员申请开通。

### (可选) 开通 DLI 服务 Spark 生命周期白名单

如果待迁移的元数据有生命周期表（即DDL带有LIFECYCLE字段），请联系DLI服务的技术支持人员申请开通Spark生命周期特性白名单。

### (可选) 开通 100 网段白名单

如果您是使用**专线**迁移方式进行数据迁移，需要提前向VPC服务申请开通“100.100.x.x”网段的白名单。

请提交工单到VPC服务，并提供以下信息进行开通：

- 华为云**账号名**、目的端DLI服务所在Region的**项目ID**：获取方法请参考[API凭证](#)。
- DLI服务的**租户名**以及租户的**项目ID**：请联系DLI服务的技术支持人员获取。

## 1.11.4 创建源端连接

创建源端连接是为了允许MgC Agent（原Edge）访问和读取源端阿里云 MaxCompute 中存储的数据。

### 前提条件

已完成[准备工作](#)。

### 操作步骤

- 步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择在完成准备工作时创建的[大数据迁移项目](#)。
- 步骤2** 在左侧导航栏选择“[实施>大数据数据迁移](#)”，进入大数据数据迁移页面。
- 步骤3** 单击页面右上角的“[连接管理](#)”按钮，进入迁移准备页面。



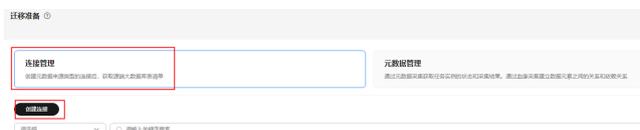
- 步骤4** 首次使用大数据数据迁移功能时，需要先启用MgC Agent。单击页面左上角的“[修改](#)”或单击“[选择迁移中心Agent](#)”按钮，弹出选择迁移工具窗口。在下拉列表中，选择在完成[准备工作](#)时与迁移中心连接成功的MgC Agent。

#### 注意

在进行大数据数据迁移时，请确保MgC Agent一直处于在线并启用状态。

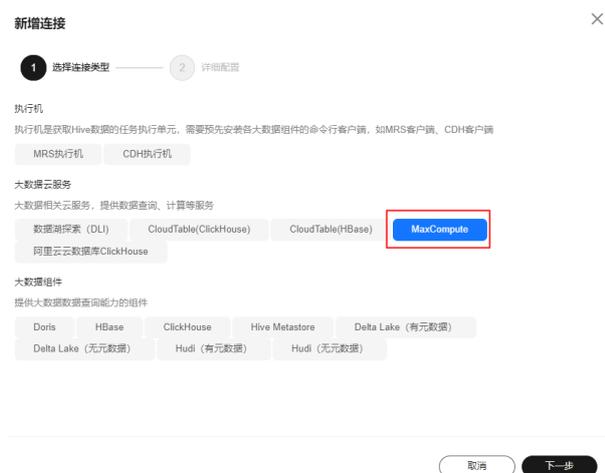
- 步骤5** 启用MgC Agent后，单击“[创建连接](#)”，右侧弹出[新增连接](#)窗口。

图 1-53 创建连接



- 步骤6** 连接类型选择“[MaxCompute](#)”，单击“[下一步](#)”。

图 1-54 选择 MaxCompute



**步骤7** 根据**MaxCompute连接参数配置说明**，配置连接参数后，单击“**测试连接**”按钮，测试通过代表连接成功。

表 1-65 MaxCompute 连接参数配置说明

| 参数        | 配置说明                                                                                                |
|-----------|-----------------------------------------------------------------------------------------------------|
| 所属源       | 选择源端。                                                                                               |
| 连接名称      | 默认创建名称为“MaxCompute-4位随机数（字母和数字）”，也可以自定义名称。                                                          |
| 迁移中心Agent | 选择在完成 <b>准备工作</b> 时与迁移中心连接成功的MgC Agent。                                                             |
| 阿里云凭证     | 选择完成 <b>准备工作</b> 时，在MgC Agent侧添加的 <b>阿里云（AK/SK）</b> 凭证。                                             |
| 项目名称      | 输入待迁移数据所在的MaxCompute项目名称。项目名称可在MaxCompute控制台的项目管理中获取。                                               |
| Endpoint  | 输入MaxCompute服务所在地域的 <b>VPC网络Endpoint</b> 信息。<br>各地域 <b>VPC网络Endpoint</b> 信息请参考 <b>Endpoint对照表</b> 。 |

创建源端连接完成后，在**连接管理**页面，可以查看已创建的连接列表。如有需要，可以单击列表操作列的“**修改链接**”对连接配置进行修改。

----结束

### 1.11.5 创建目的端连接

创建目的端连接是为了将源端数据写入到华为云DLI。

## 使用须知

**元数据迁移：**只需要配置一条连接到DLI SQL队列的连接。

**数据迁移：**需要配置两条连接，分别为连接到DLI SQL队列的连接和连接到DLI通用队列的连接。

## 前提条件

已完成**准备工作**。

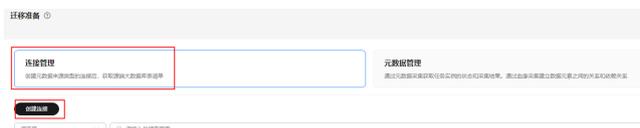
## 操作步骤

- 步骤1** 登录**迁移中心管理控制台**。在左侧导航栏的项目下拉列表中选择在完成准备工作时创建的**大数据迁移项目**。
- 步骤2** 在左侧导航栏选择“**实施>大数据数据迁移**”，进入**大数据数据迁移**页面。
- 步骤3** 单击页面右上角的“**连接管理**”按钮，进入**迁移准备**页面。



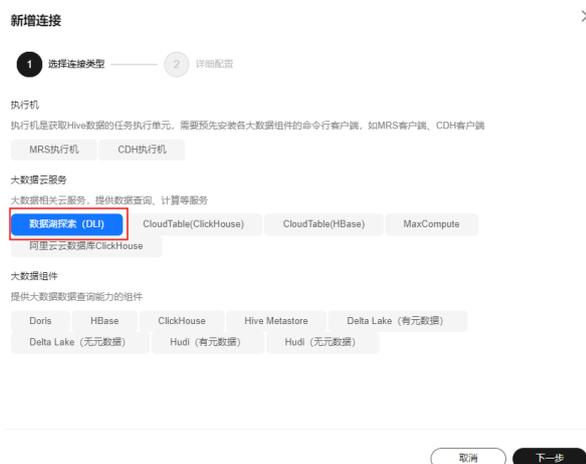
- 步骤4** 单击“**创建连接**”，右侧弹出**新增连接**窗口。

图 1-55 创建连接



- 步骤5** 连接类型选择“**数据湖探索（DLI）**”，单击“**下一步**”，进入**详细配置**页面。

图 1-56 选择 DLI



**步骤6** 根据[DLI服务参数配置说明](#)，配置连接参数后，单击“**测试连接**”按钮，测试通过代表连接成功。

**表 1-66** DLI 服务参数配置说明

| 参数        | 配置说明                                                                                                                                                                                               |
|-----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 所属源       | 选择目的端。                                                                                                                                                                                             |
| 连接名称      | 默认创建名称为“DLI-4位随机数（字母和数字）”，也可以自定义名称。                                                                                                                                                                |
| 迁移中心Agent | 选择在完成 <a href="#">准备工作</a> 时与迁移中心连接成功的MgC Agent。                                                                                                                                                   |
| DLI凭证     | 选择完成 <a href="#">准备工作</a> 时，在MgC Agent侧添加的 <b>华为云（AK/SK）</b> 凭证。如果所选凭证与登录迁移中心账号凭证相同，请勾选“ <b>所选凭证与登录迁移中心凭证相同</b> ”，系统将根据所选的区域，自动查询账号在该区域下的项目。                                                       |
| 区域        | 选择目的端DLI所在区域。如果未勾选“ <b>所选凭证与登录迁移中心凭证相同</b> ”，则需要手动填写区域代号。例如：cn-south-1                                                                                                                             |
| 项目        | 选择DLI在目的端区域用于接收数据的项目。如果未勾选“ <b>所选凭证与登录迁移中心凭证相同</b> ”，则需要手动填写项目ID。项目ID的获取方法请参考 <a href="#">获取项目的名称和ID</a> 。<br> |
| 队列名称      | 输入完成 <a href="#">准备工作</a> 时，创建的SQL队列名称/通用队列名称。 <ul style="list-style-type: none"> <li>当配置SQL队列连接时，填写SQL队列名称。</li> <li>当配置通用队列连接时，填写通用队列名称。</li> </ul>                                              |

创建目的端连接完成后，在[连接管理](#)页面，可以查看已创建的连接列表。如有需要，可以单击列表操作列的“**修改链接**”对连接配置进行修改。

---结束

## 1.11.6 创建迁移任务

### 1.11.6.1 元数据全量迁移

将源端数据库中的所有元数据迁移到华为云数据湖探索（DLI）。

#### 须知

- 元数据迁移不支持事务表（即DDL中带有transactional=true字段的表）。
- 当迁移的表数量较多或单个表的列数较多时，迁移过程可能会耗费较长时间。

## 前提条件

- 已完成所有准备工作。
- 已创建源端连接。
- 已创建目的端连接。

## 操作步骤

**步骤1** 登录**迁移中心管理控制台**。在左侧导航栏的项目下拉列表中选择在完成准备工作时创建的**大数据迁移项目**。

**步骤2** 在左侧导航栏选择“**实施>大数据数据迁移**”，进入**大数据数据迁移**页面。

**步骤3** 单击页面右上角的“**创建迁移任务**”按钮，右侧弹出**创建迁移任务**窗口。



**步骤4** 源端组件选择“**MaxCompute**”，目的端组件选择“**数据湖探索 DLI**”，任务类型选择“**元数据全量迁移**”，单击“**下一步**”，进入**详细配置**页面。



**步骤5** 根据**表1-67**，配置元数据全量迁移任务参数。

**表 1-67** 元数据全量迁移参数配置说明

| 区域   | 参数        | 配置说明                                                  |
|------|-----------|-------------------------------------------------------|
| 基本信息 | 任务名称      | 默认创建名称为“MaxCompute到DLI元数据全量迁移-4位随机数（字母和数字）”，也可以自定义名称。 |
|      | 迁移中心Agent | 选择在完成 <b>准备工作</b> 时与迁移中心连接成功的MgC Agent。               |

| 区域   | 参数           | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|------|--------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 源端配置 | 源端连接         | 选择 <a href="#">创建的源端连接</a> 。                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|      | MaxCompute参数 | 非必填参数，一般为空。如有需要，可参考 <a href="#">MaxCompute参数文档</a> 进行配置。                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| 数据范围 | 按库筛选         | <p>即按数据库进行筛选，在数据库参数框中输入待迁移的数据库名称。单击“添加数据库”，可以新增数据库，最多支持同时添加10个数据库。</p>  <p>如果数据库中有不需要迁移的表，可以下载CSV格式的“排除表”模板，参考<b>b.~e.</b>在模板中添加不需要迁移的表信息，然后上传至MgC。</p>                                                                                                                                                                                                                                                              |
|      | 按表筛选         | <ol style="list-style-type: none"> <li>1. 下载CSV格式的模板。</li> <li>2. 下载后用<b>记事本</b>打开该CSV模板文件。<br/><b>注意</b><br/>请不要使用Excel对CSV模板文件进行编辑，通过Excel编辑并保存的模板文件无法被MgC识别。</li> <li>3. 保留CSV文件中的第一行不变，从第二行开始，每行填写一个表的迁移信息，格式为“{MC项目名},{表名}”，其中MC项目名代表MaxCompute项目名称，表名代表具体的表名称。<br/><b>须知</b> <ul style="list-style-type: none"> <li>- 每行的“MC项目名”和“表名”之间使用<b>英文逗号</b>隔开，不要使用空格或其他分隔符。</li> <li>- 每添加完一个表的信息后，使用回车键换行。</li> </ul> </li> <li>4. 完成所有表信息的添加后，保存CSV文件的更改。</li> <li>5. 将编辑并保存好的CSV文件上传至MgC。</li> </ol> |
|      | 企业项目ID（可选）   | 输入目的端DLI数据库所属的企业项目ID。企业项目ID的获取方法请参考 <a href="#">如何获取企业项目ID</a> 。                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|      | 内表转换         | <ul style="list-style-type: none"> <li>● 选择内表进行迁移时，迁移后的数据会存储在DLI中。</li> <li>● 选择外表进行迁移时，迁移后的数据会存储在您指定的OBS路径下。</li> </ul>                                                                                                                                                                                                                                                                                                                                                                               |

| 区域    | 参数        | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|-------|-----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 目的端配置 | 覆盖策略      | 对已有表的处理策略 <ul style="list-style-type: none"> <li>• 跳过：如果目的端已经存在同名的数据表，将忽略该表，不会对其进行任何操作。</li> <li>• 重建：如果目的端已经存在同名的数据表，该表及其数据将被完全删除，并根据源端的定义重新创建。</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|       | 目的端连接     | 选择创建的 <b>目的端DLI SQL队列连接</b> 。<br><b>注意</b><br>切勿选择通用队列。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|       | 自定义参数     | 根据实际需求进行配置，支持的自定义参数详情请参考 <b>conf参数说明</b> 和 <b>自定义参数说明</b> 。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|       | 表类型       | 根据实际情况选择。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|       | 文件类型      | 根据实际情况选择。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|       | 压缩格式      | 根据实际情况选择。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| 迁移选项  | 并发连接数     | 设置迁移并发数。默认为3，1≤取值范围≤10。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|       | 表路径映射（可选） | <p>用于确保在元数据迁移过程中，MaxCompute外表的元数据能够正确地迁移到目的端指定的存储位置，即在OSS中的源路径自动映射为目的端指定的OBS路径。</p> <p>映射到目的端OBS的路径需要提前创建，并且除桶名以外，其余目录需要与源端OSS目录保存一致。</p> <p>例如，源端表所在目录路径为：<code>oss://oss-cn-hangzhou.aliyuncs.com/ali-test/documents/example_table</code>，其中<code>oss-cn-hangzhou.aliyuncs.com</code>为桶所在地域的Endpoint，<code>ali-test</code>为桶名，<code>documents</code>为表的根目录，<code>example_table</code>为表目录。那么，映射到目的端的路径必须为：<code>obs://hw-test/documents/example_table</code>，除桶名外，根目录和表目录必须和源端一致。</p> <p>填写示例：</p> <ul style="list-style-type: none"> <li>• 源端：<code>oss://oss-cn-hangzhou.aliyuncs.com/ali-test/documents/example_table</code></li> <li>• 目的端：<code>obs://hw-test/documents/example_table</code></li> </ul> <p><b>说明</b><br/>以上路径仅为示例，请在实际使用中替换为正确的路径。</p> |

**步骤6** 配置完成后，选择任务的运行方式。

## 须知

- 创建的迁移任务可重复运行，每运行一次，即会生成一条任务实例。
- 单击任务名称，可以修改任务配置。

- **立即运行**：如果需要立即运行迁移任务，可以勾选“**保存后立即运行**”，然后单击“**保存**”按钮，会在**任务管理**列表中生成一条元数据全量迁移任务并自动开始执行迁移。

迁移选项

并发连接数

3

请输入区间为1-10的并发连接数

表路径映射 (可选)

请补充源端表路径到目的端表路径的映射关系，该映射关系将会在目的端创建的DDL中替换掉相应表路径字段。

查看示例

请输入oss桶路径

请输入obs桶路径

保存后立即运行

取消

上一步

保存

- **手动运行**：如果需要后续手动启动迁移任务，则只需单击“**保存**”按钮，在**任务管理**列表中生成一条元数据全量迁移任务。在需要启动迁移时，单击操作列的“**运行**”，手动启动迁移。



**步骤7** 迁移任务运行后，单击操作列的“**查看实例**”，进入**任务进度**页签，可以查看该任务产生的运行实例以及迁移进度。



单击进度统计列的“**查看**”，进入进度详情页面，可以[查看并导出任务结果](#)。

----结束

## 1.11.6.2 元数据增量感知

识别出上一次迁移后，在源端数据库中发生变化的元数据，并同步至目的端华为云数据湖探索（DLI）。

### 前提条件

- 已**创建源端连接**。
- 已**创建目的端连接**。
- 最少完成一次**元数据全量迁移**。

## 准备工作

- **开通spark 3.3.1白名单**  
使用该功能前，请联系技术支持人员，将您添加到spark 3.3.1版本的白名单中。
- **配置DLI作业桶**  
需要购买一个OBS桶或并行文件系统并设置为DLI作业桶，该桶用于存储DLI作业运行过程中产生的临时数据。方法请参考[配置DLI作业桶](#)。

## 操作步骤

- 步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择在完成准备工作时创建的[大数据迁移项目](#)。
- 步骤2** 在左侧导航栏选择“[实施>大数据数据迁移](#)”，进入[大数据数据迁移](#)页面。
- 步骤3** 单击页面右上角的“[创建迁移任务](#)”按钮，右侧弹出[创建迁移任务](#)窗口。



- 步骤4** 源端组件选择“[MaxCompute](#)”，目的端组件选择“[数据湖探索 DLI](#)”，任务类型选择“[元数据增量感知](#)”，单击“[下一步](#)”，进入[详细配置](#)页面。



- 步骤5** 根据[表1-68](#)，配置元数据增量感知任务参数。

**表 1-68** 元数据增量感知参数配置说明

| 区域   | 参数   | 配置说明                                                  |
|------|------|-------------------------------------------------------|
| 基本信息 | 任务名称 | 默认创建名称为“MaxCompute到DLI元数据增量感知-4位随机数（字母和数字）”，也可以自定义名称。 |

| 区域   | 参数               | 配置说明                                                                                                                                                                                                                                                                                                                    |
|------|------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|      | 迁移中心Agent        | 选择在完成准备工作时与迁移中心连接成功的MgC Agent。                                                                                                                                                                                                                                                                                          |
| 源端配置 | 源端连接             | 选择创建的源端连接。                                                                                                                                                                                                                                                                                                              |
|      | MaxCompute参数（可选） | 非必填参数，一般为空。如有需要可参考 <a href="#">MaxCompute参数文档</a> 进行配置。                                                                                                                                                                                                                                                                 |
| 数据范围 | 时间范围             | <ul style="list-style-type: none"> <li>全量区间：全量识别出上一次元数据迁移后，在源端数据库中发生变化的元数据。</li> <li>增量区间：选择T-N区间作为增量元数据的识别范围，即任务开始执行时间的前 24 x N 小时内的增量元数据。例如：选择T-1，2024/06/06 14:50执行任务，则表示识别从2024/06/05 14:50 到 2024/06/06 14:50（现在）的增量元数据。</li> </ul> <p><b>须知</b><br/>增量区间任务仅会识别并同步在指定时间范围内发生变化的元数据。仅源端存在和仅目的端存在的元数据，都不会被同步或处理。</p> |
|      | 按库筛选             | <p>即按数据库进行筛选，在数据库参数框中输入需要增量待迁移的数据库名称。单击“添加数据库”，可以新增数据库，最多支持同时添加10个数据库。</p>  <p>如果数据库中有不需要迁移的表，可以下载CSV格式的“排除表”模板，参考<b>b~e</b>。在模板中添加不需要迁移的表信息，然后上传至MgC。</p>                                                                          |

| 区域    | 参数        | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|-------|-----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|       | 按表筛选      | <ol style="list-style-type: none"> <li>1. 下载CSV格式的模板。</li> <li>2. 下载后用<b>记事本</b>打开该CSV模板文件。<br/><b>注意</b><br/>请不要使用Excel对CSV模板文件进行编辑，通过Excel编辑并保存的模板文件无法被MgC识别。</li> <li>3. 保留CSV文件中的第一行不变，从第二行开始，每行填写一个表的迁移信息，格式为“{MC项目名},{表名}”，其中MC项目名代表MaxCompute项目名称，表名代表具体的表名称。<br/><b>须知</b> <ul style="list-style-type: none"> <li>- 每行的“MC项目名”和“表名”之间使用<b>英文逗号</b>隔开，不要使用空格或其他分隔符。</li> <li>- 每添加完一个表的信息后，使用回车键换行。</li> </ul> </li> <li>4. 完成所有表信息的添加后，保存CSV文件的更改。</li> <li>5. 将编辑并保存好的CSV文件上传至MgC。</li> </ol> |
| 目的端配置 | 目的端连接     | 选择创建的 <b>目的端DLI SQL队列连接</b> 。<br><b>注意</b><br>切勿选择通用队列。                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|       | 自定义参数（可选） | 根据实际需求进行配置，支持的自定义参数详情请参考 <a href="#">conf参数说明</a> 和 <a href="#">自定义参数说明</a> 。                                                                                                                                                                                                                                                                                                                                                                                                                          |
| 迁移选项  | 并发连接数     | 设置迁移并发数。默认为3，1≤取值范围≤10。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |

**步骤6** 配置完成后，选择任务的运行方式。

**须知**

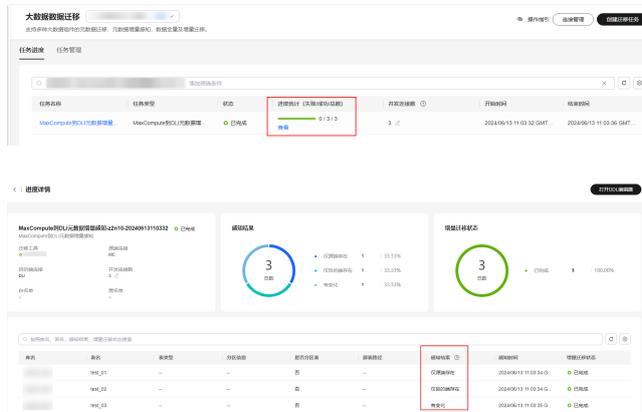
- 创建的迁移任务可重复运行，每运行一次，即会生成一条任务实例。
  - 单击任务名称，可以修改任务配置。
- 
- **立即运行**：如果需要立即运行迁移任务，可以勾选“**保存后立即运行**”，然后单击“**保存**”按钮，会在**任务管理**列表中生成一条元数据增量感知任务并自动开始执行迁移。
- The screenshot shows a configuration form for migration options. Under the heading '迁移选项' (Migration Options), there is a field for '并发连接数' (Concurrent Connections) with a value of 3 and a range of 1-10. Below this field, there is a checkbox labeled '保存后立即运行' (Save and Run Immediately) which is checked. To the right of the checkbox are buttons for '取消' (Cancel), '上一步' (Previous Step), and '保存' (Save). The '保存' button is highlighted with a red box.
- **手动运行**：如果需要后续手动启动迁移任务，则只需单击“**保存**”按钮，在**任务管理**列表中生成一条元数据增量感知任务。在需要启动迁移时，单击操作列的“**运行**”，手动启动迁移。



**步骤7** 迁移任务运行后，单击操作列的“查看实例”，进入任务进度页签，可以查看该任务产生的运行实例以及迁移进度。



单击进度统计列的“查看”，进入进度详情页面。可以查看元数据的增量感知结果并导出任务结果。



**步骤8** 单击进度详情页面右上角的“打开DDL编辑器”，可以根据不同的增量感知结果，对源端和目的端的表结构进行比较和编辑。



----结束

### 1.11.6.3 全量数据迁移

将源端数据库中的所有数据迁移到目的端华为云数据湖探索（DLI）。

#### 使用须知

- 当迁移的表数量较多或单个表的列数较多时，迁移过程可能会耗费较长时间。
- 在数据迁移时，如果重复迁移分区数据时提示错误信息：“Do not support update partition SerDeInfo for DLI table”，需要您手动删除DLI中的分区数据，然后再进行迁移。

## 前提条件

- 已完成所有准备工作。
- 已创建源端连接。
- 已创建目的端连接。
- 开始迁移前，需要联系技术支持人员，将您添加到允许jar程序访问DLI元数据功能的白名单中。

## 操作步骤

**步骤1** 登录**迁移中心管理控制台**。在左侧导航栏的项目下拉列表中选择在完成准备工作时创建的大数据迁移项目。

**步骤2** 在左侧导航栏选择“**实施>大数据数据迁移**”，进入大数据数据迁移页面。

**步骤3** 单击页面右上角的“**创建迁移任务**”按钮，右侧弹出创建迁移任务窗口。



**步骤4** 源端组件选择“**MaxCompute**”，目的端组件选择“**数据湖探索 DLI**”，任务类型选择“**全量数据迁移**”，单击“**下一步**”，进入详细配置页面。



**步骤5** 根据**表1-69**，配置全量数据迁移任务参数。

**表 1-69** 全量数据迁移参数配置说明

| 区域   | 参数   | 配置说明                                                 |
|------|------|------------------------------------------------------|
| 基本信息 | 任务名称 | 默认创建名称为“MaxCompute到DLI全量数据迁移-4位随机数（字母和数字）”，也可以自定义名称。 |

| 区域    | 参数               | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|-------|------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|       | 迁移中心Agent        | 选择在完成准备工作时与迁移中心连接成功的MgC Agent。                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| 源端配置  | 源端连接             | 选择创建的源端连接。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|       | MaxCompute参数（可选） | 非必填参数，一般为空。如有需要，可参考 <a href="#">MaxCompute参数文档</a> 进行配置。                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| 数据范围  | 预估项目周期（天）（可选）    | 配置后，系统会在任务执行时检查表的生命周期。若表的生命周期在项目的预期结束时间之前终止，则忽略该表。不配置则默认迁移所有表。                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|       | 按库筛选             | <p>即按数据库进行筛选，在数据库参数框中输入待迁移的数据库名称。单击“添加数据库”，可以新增数据库，最多支持同时添加10个数据库。</p>  <p>如果数据库中有不需要迁移的表，可以下载CSV格式的“排除表”模板，参考<b>b.~e.</b>在模板中添加不需要迁移的表信息，然后上传至MgC。</p>                                                                                                                                                                                                                                                      |
|       | 按表筛选             | <ol style="list-style-type: none"> <li>1. 下载CSV格式的模板。</li> <li>2. 下载后用记事本打开该CSV模板文件。<br/><b>注意</b><br/>请不要使用Excel对CSV模板文件进行编辑，通过Excel编辑并保存的模板文件无法被MgC识别。</li> <li>3. 保留CSV文件中的第一行不变，从第二行开始，每行填写一个表的迁移信息，格式为“{MC项目名},{表名}”，其中MC项目名代表MaxCompute项目名称，表名代表具体的表名称。<br/><b>须知</b> <ul style="list-style-type: none"> <li>- 每行的“MC项目名”和“表名”之间使用<b>英文逗号</b>隔开，不要使用空格或其他分隔符。</li> <li>- 每添加完一个表的信息后，使用回车键换行。</li> </ul> </li> <li>4. 完成所有表信息的添加后，保存CSV文件的更改。</li> <li>5. 将编辑并保存好的CSV文件上传至MgC。</li> </ol> |
| 目的端配置 | 目的端连接            | <p>选择创建的<b>目的端DLI 通用队列连接</b>。</p> <p><b>注意</b><br/>切勿选择DLI SQL队列。</p>                                                                                                                                                                                                                                                                                                                                                                                                                           |

| 区域                                       | 参数                                          | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                                         |        |   |                                        |        |   |                             |                                  |   |                                    |                                   |   |
|------------------------------------------|---------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------|------|---|---------------------------|--|---|-----------------------------------------|------------------------------------------|---|------------------------------------------|---------------------------------------------|---|-----------------------------|------|---|---------------------------|--|---|-----------------------------------------|------------------------------------------|---|------------------------------------------|---------------------------------------------|---|-----------------------------------------|--------|---|----------------------------------------|--------|---|-----------------------------|----------------------------------|---|------------------------------------|-----------------------------------|---|
|                                          | 自定义参数（可选）                                   | <p>根据实际需求进行配置，支持的自定义参数详情请参考<a href="#">conf参数说明</a>和<a href="#">自定义参数说明</a>。</p> <ul style="list-style-type: none"> <li>如果是<b>公网迁移</b>，需要填写以下4个参数：           <div data-bbox="858 421 1372 631" style="border: 1px solid #ccc; padding: 5px; margin: 5px 0;"> <p>自定义参数<br/>参数详情请参考文档中“表3 conf参数说明”部分。 <a href="#">查看文档</a></p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 60%;">spark.dli.metaAccess.enable</td> <td>true</td> <td style="text-align: right;">✕</td> </tr> <tr> <td>spark.dli.job.agency.name</td> <td></td> <td style="text-align: right;">✕</td> </tr> <tr> <td>mgc.mc2dli.data.migration.dli.file.path</td> <td>obs://mgc-test/data/migration-dli-spark-</td> <td style="text-align: right;">✕</td> </tr> <tr> <td>mgc.mc2dli.data.migration.dli.spark.jars</td> <td>["obs://mgc-test/data/datasource.jar","oi"]</td> <td style="text-align: right;">✕</td> </tr> </table> <p style="text-align: center; color: #00a0e3; font-size: small;">添加参数</p> </div> <ul style="list-style-type: none"> <li>- spark.dli.metaAccess.enable：填写true。</li> <li>- spark.dli.job.agency.name：填写在完成准备工作时配置的<b>DLI委托权限名称</b>。</li> <li>- mgc.mc2dli.data.migration.dli.file.path：填写在完成准备工作时<b>存放migration-dli-spark-1.0.0.jar包的OBS路径</b>。例如：<br/>obs://mgc-test/data/migration-dli-spark-1.0.0.jar</li> <li>- mgc.mc2dli.data.migration.dli.spark.jars：填写在完成准备工作时<b>存放fastjson-1.2.54.jar和datasource.jar包的OBS路径</b>，以数组形式传入，路径带双引号，以英文逗号隔开。例如：["obs://mgc-test/data/datasource.jar","obs://mgc-test/data/fastjson-1.2.54.jar"]</li> </ul> <li>如果是<b>专线迁移</b>，需要填写以下8个参数：           <div data-bbox="858 1236 1372 1601" style="border: 1px solid #ccc; padding: 5px; margin: 5px 0;"> <p>自定义参数<br/>参数详情请参考文档中“表3 conf参数说明”部分。 <a href="#">查看文档</a></p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 60%;">spark.dli.metaAccess.enable</td> <td>true</td> <td style="text-align: right;">✕</td> </tr> <tr> <td>spark.dli.job.agency.name</td> <td></td> <td style="text-align: right;">✕</td> </tr> <tr> <td>mgc.mc2dli.data.migration.dli.file.path</td> <td>obs://mgc-test/data/migration-dli-spark-</td> <td style="text-align: right;">✕</td> </tr> <tr> <td>mgc.mc2dli.data.migration.dli.spark.jars</td> <td>["obs://mgc-test/data/datasource.jar","oi"]</td> <td style="text-align: right;">✕</td> </tr> <tr> <td>spark.sql.catalog.mc_catalog.tableWrite</td> <td>tunnel</td> <td style="text-align: right;">✕</td> </tr> <tr> <td>spark.sql.catalog.mc_catalog.tableReac</td> <td>tunnel</td> <td style="text-align: right;">✕</td> </tr> <tr> <td>spark.hadoop.odps.end.point</td> <td>http://service.ci.***.maxcompute</td> <td style="text-align: right;">✕</td> </tr> <tr> <td>spark.hadoop.odps.tunnel.end.point</td> <td>http://dt.ci.***.maxcompute.aliyt</td> <td style="text-align: right;">✕</td> </tr> </table> <p style="text-align: center; color: #00a0e3; font-size: small;">添加参数</p> </div> <ul style="list-style-type: none"> <li>- spark.dli.metaAccess.enable：填写true。</li> <li>- spark.dli.job.agency.name：填写在完成准备工作时配置的<b>DLI委托权限名称</b>。</li> <li>- mgc.mc2dli.data.migration.dli.file.path：填写在完成准备工作时<b>存放migration-dli-spark-1.0.0.jar包的OBS路径</b>。例如：<br/>obs://mgc-test/data/migration-dli-spark-1.0.0.jar</li> <li>- mgc.mc2dli.data.migration.dli.spark.jars：填写在完成准备工作时<b>存放</b></li> </ul> </li> </li></ul> | spark.dli.metaAccess.enable | true | ✕ | spark.dli.job.agency.name |  | ✕ | mgc.mc2dli.data.migration.dli.file.path | obs://mgc-test/data/migration-dli-spark- | ✕ | mgc.mc2dli.data.migration.dli.spark.jars | ["obs://mgc-test/data/datasource.jar","oi"] | ✕ | spark.dli.metaAccess.enable | true | ✕ | spark.dli.job.agency.name |  | ✕ | mgc.mc2dli.data.migration.dli.file.path | obs://mgc-test/data/migration-dli-spark- | ✕ | mgc.mc2dli.data.migration.dli.spark.jars | ["obs://mgc-test/data/datasource.jar","oi"] | ✕ | spark.sql.catalog.mc_catalog.tableWrite | tunnel | ✕ | spark.sql.catalog.mc_catalog.tableReac | tunnel | ✕ | spark.hadoop.odps.end.point | http://service.ci.***.maxcompute | ✕ | spark.hadoop.odps.tunnel.end.point | http://dt.ci.***.maxcompute.aliyt | ✕ |
| spark.dli.metaAccess.enable              | true                                        | ✕                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                                         |        |   |                                        |        |   |                             |                                  |   |                                    |                                   |   |
| spark.dli.job.agency.name                |                                             | ✕                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                                         |        |   |                                        |        |   |                             |                                  |   |                                    |                                   |   |
| mgc.mc2dli.data.migration.dli.file.path  | obs://mgc-test/data/migration-dli-spark-    | ✕                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                                         |        |   |                                        |        |   |                             |                                  |   |                                    |                                   |   |
| mgc.mc2dli.data.migration.dli.spark.jars | ["obs://mgc-test/data/datasource.jar","oi"] | ✕                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                                         |        |   |                                        |        |   |                             |                                  |   |                                    |                                   |   |
| spark.dli.metaAccess.enable              | true                                        | ✕                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                                         |        |   |                                        |        |   |                             |                                  |   |                                    |                                   |   |
| spark.dli.job.agency.name                |                                             | ✕                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                                         |        |   |                                        |        |   |                             |                                  |   |                                    |                                   |   |
| mgc.mc2dli.data.migration.dli.file.path  | obs://mgc-test/data/migration-dli-spark-    | ✕                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                                         |        |   |                                        |        |   |                             |                                  |   |                                    |                                   |   |
| mgc.mc2dli.data.migration.dli.spark.jars | ["obs://mgc-test/data/datasource.jar","oi"] | ✕                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                                         |        |   |                                        |        |   |                             |                                  |   |                                    |                                   |   |
| spark.sql.catalog.mc_catalog.tableWrite  | tunnel                                      | ✕                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                                         |        |   |                                        |        |   |                             |                                  |   |                                    |                                   |   |
| spark.sql.catalog.mc_catalog.tableReac   | tunnel                                      | ✕                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                                         |        |   |                                        |        |   |                             |                                  |   |                                    |                                   |   |
| spark.hadoop.odps.end.point              | http://service.ci.***.maxcompute            | ✕                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                                         |        |   |                                        |        |   |                             |                                  |   |                                    |                                   |   |
| spark.hadoop.odps.tunnel.end.point       | http://dt.ci.***.maxcompute.aliyt           | ✕                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                                         |        |   |                                        |        |   |                             |                                  |   |                                    |                                   |   |

| 区域   | 参数      | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|------|---------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|      |         | <p><b>fastjson-1.2.54.jar和datasource.jar包的OBS路径</b>，以数组形式传入，路径带双引号，以英文逗号隔开。例如：["obs://mgc-test/data/datasource.jar","obs://mgc-test/data/fastjson-1.2.54.jar"]</p> <ul style="list-style-type: none"> <li>- spark.sql.catalog.mc_catalog.tableWriteProvider: 填写tunnel。</li> <li>- spark.sql.catalog.mc_catalog.tableReadProvider: 填写tunnel。</li> <li>- spark.hadoop.odps.end.point: 填写源端MaxCompute服务所在地域的<b>VPC网络Endpoint</b>信息。各地域<b>VPC网络Endpoint</b>信息请参考<a href="#">Endpoint对照表</a>。例如：源端所在地域为“香港”，则填写“http://service.cn-hongkong.maxcompute.aliyun-inc.com/api”</li> <li>- spark.hadoop.odps.tunnel.end.point: 填写源端MaxCompute服务所在地域的<b>VPC网络Tunnel Endpoint</b>信息。各地域<b>VPC网络Tunnel Endpoint</b>信息请参考<a href="#">Endpoint对照表</a>。例如：源端所在地域为“香港”，则填写“http://dt.cn-hongkong.maxcompute.aliyun-inc.com”</li> </ul> |
| 迁移选项 | 大表规则    | 设置大表条件，当符合大表的条件时，将自动按分区拆分迁移子任务。建议保持默认，也可根据实际需求修改。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|      | 小表规则    | 设置小表条件，当符合小表的条件时，自动将多个小表合并到一个迁移子任务中，提高迁移效率。建议保持默认，也可根据实际需求修改。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|      | 并发连接数   | 设置迁移并发数。默认为3，1≤取值范围≤10。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|      | 单文件SQL数 | 每批提交的文件中所包含的SQL条数。默认为3，1≤推荐取值范围≤50。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |

**步骤6** 配置完成后，选择任务的运行方式。

**须知**

- 创建的迁移任务可重复运行，每运行一次，即会生成一条任务实例。
- 单击任务名称，可以修改任务配置。
- **立即运行**：如果需要立即运行迁移任务，可以勾选“**保存后立即运行**”，然后单击“**保存**”按钮，会在**任务管理**列表中生成一条全量数据迁移任务并自动开始执行迁移。



- **手动运行**：如果需要后续手动启动迁移任务，则只需单击“保存”按钮，在任务管理列表中生成一条全量数量迁移任务。在需要启动迁移时，单击操作列的“运行”，手动启动迁移。



**步骤7** 迁移任务运行后，单击操作列的“查看实例”，进入任务进度页签，可以查看该任务产生的运行实例以及迁移进度。



单击状态列的“重新运行”，可以多次运行迁移实例。



单击进度统计列的“查看”，进入进度详情页面。可以查看迁移的数据表清单、状态并导出任务结果。



**步骤8** （可选）数据迁移完成后，可以使用MgC提供的大数据校验功能，对源端和目的端数据进行一致性校验。详细的操作步骤请参考[MaxCompute迁移至DLI数据校验](#)。

----结束

### 1.11.6.4 增量数据迁移

将上一次迁移后，在源端数据库中发生变化（新增、修改或删除）的数据同步至目的端华为云数据湖探索（DLI）。

#### 使用须知

在进行指定分区数据迁移的场景下，由于MaxCompute在创建表时会自动将表名转换为小写，因此在迁移过程中，请确保Excel文件中的库表名称与MaxCompute中的实际表名一致（全部小写）。如果Excel文件中填写的库表名称为大写，迁移任务会因为无法正确识别表名导致无法迁移。

#### 前提条件

- 已完成所有准备工作。
- 已创建源端连接。
- 已创建目的端连接。
- 已最少完成一次全量数据迁移。
- 开始迁移前，需要联系技术支持人员，将您添加到允许jar程序访问DLI元数据功能的白名单中。

#### 操作步骤

**步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择在完成准备工作时创建的大数据迁移项目。

**步骤2** 在左侧导航栏选择“[实施>大数据数据迁移](#)”，进入大数据数据迁移页面。

**步骤3** 单击页面右上角的“[创建迁移任务](#)”按钮，右侧弹出创建迁移任务窗口。



**步骤4** 源端组件选择“[MaxCompute](#)”，目的端组件选择“[数据湖探索 DLI](#)”，任务类型选择“[增量数据迁移](#)”，单击“[下一步](#)”，进入详细配置页面。



**步骤5** 根据表1-70，配置增量数据迁移任务参数。

**表 1-70** 增量数据迁移参数配置说明

| 区域   | 参数               | 配置说明                                                 |
|------|------------------|------------------------------------------------------|
| 基本信息 | 任务名称             | 默认创建名称为“MaxCompute到DLI增量数据迁移-4位随机数（字母和数字）”，也可以自定义名称。 |
|      | 迁移中心Agent        | 选择在完成准备工作时与迁移中心连接成功的MgC Agent。                       |
| 源端配置 | 源端连接             | 选择创建的源端连接。                                           |
|      | MaxCompute参数（可选） | 非必填参数，一般为空。如有需要，可参考MaxCompute参数文档进行配置。               |

| 区域   | 参数     | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|------|--------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 数据范围 | 数据迁移策略 | <ul style="list-style-type: none"> <li>● <b>时间过滤</b>: 通过数据变化发生的时间, 过滤出需要迁移的增量数据。选择该方式还需要配置“增量区间”、“分区筛选方式”、“按库筛选或按表筛选”等参数。</li> <li>● <b>自定义过滤</b>: 迁移指定分区内发生变化的数据。选择该方式, 请按照以下步骤, 在模板文件中添加分区表信息, 然后上传至MgC。               <ol style="list-style-type: none"> <li>1. 单击“指定分区迁移文件”下的“下载模板”, 下载CSV格式的模板。</li> <li>2. 下载后用<b>记事本</b>打开该CSV模板文件。                   <p><b>注意</b><br/>请不要使用Excel对CSV模板文件进行编辑, 通过Excel编辑并保存的模板文件无法被MgC识别。</p> </li> <li>3. 保留CSV文件中的第一行不变, 从第二行开始, 每行填写一个表的迁移信息, 格式为“{MC项目名},{表名},{分区字段},{分区值}”, 其中MC项目名代表MaxCompute项目名称, 表名代表具体的表名称。                   <p><b>须知</b><br/>每行的参数值之间使用<b>英文逗号</b>隔开, 不要使用空格或其他分隔符。<br/>每添加完一个表的信息后, 使用回车键换行。</p> </li> <li>4. 完成所有表信息的添加后, 保存CSV文件的更改。</li> <li>5. 单击“指定分区迁移文件”下的“添加文件”, 将编辑并保存好的CSV文件上传至MgC。</li> </ol> </li> </ul> |
|      | 增量区间   | <p>选择T-N区间作为增量数据的迁移范围, 即任务开始执行时间的前 24 x N 小时内的增量数据。例如: 选择T-1, 2024/06/06 14:50执行任务, 则表示迁移从 2024/06/05 14:50 到 2024/06/06 14:50 (现在) 的增量数据。</p> <p>如果勾选“指定日期”, 将只迁移指定日期内的增量数据。</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|      | 分区筛选方式 | <p>根据数据的分区属性来选择和检索数据。默认为<b>更新时间</b>。</p> <ul style="list-style-type: none"> <li>● <b>更新时间</b>: 筛选出最近或在特定时间段内有更新的数据。</li> <li>● <b>创建时间</b>: 筛选出在特定时间窗口内创建的数据。</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |

| 区域    | 参数    | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|-------|-------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|       | 按库筛选  | <p>即按数据库进行筛选，在数据库参数框中输入需要增量迁移的数据库名称。单击“添加数据库”，可以新增数据库，最多支持同时添加10个数据库。</p>  <p>数据范围<br/>筛选方式<br/>按库筛选 按表筛选</p> <p>数据库<br/>点击“添加数据库”后，在输入框中输入数据库名</p> <p>添加数据库</p> <p>您还可以增加8个数据库。</p> <p>排除表（可选）<br/>添加文件 下载模板</p> <p>只能上传CSV文件，且不能超过1M，表格数据将在上传后自动去重。</p> <p>如果数据库中有不需要迁移的表，可以下载CSV格式的“排除表”模板，参考<b>b.~e.</b>在模板中添加不需要迁移的表信息，然后上传至MgC。</p>                                                                                                                                                                   |
|       | 按表筛选  | <ol style="list-style-type: none"> <li>1. 下载CSV格式的模板。</li> <li>2. 下载后用<b>记事本</b>打开该CSV模板文件。</li> </ol> <p><b>注意</b><br/>请不要使用Excel对CSV模板文件进行编辑，通过Excel编辑并保存的模板文件无法被MgC识别。</p> <ol style="list-style-type: none"> <li>3. 保留CSV文件中的第一行不变，从第二行开始，每行填写一个表的迁移信息，格式为“{MC项目名},{表名}”，其中MC项目名代表MaxCompute项目名称，表名代表具体的表名称。</li> </ol> <p><b>须知</b></p> <ul style="list-style-type: none"> <li>- 每行的“MC项目名”和“表名”之间使用<b>英文逗号</b>隔开，不要使用空格或其他分隔符。</li> <li>- 每添加完一个表的信息后，使用回车键换行。</li> </ul> <ol style="list-style-type: none"> <li>4. 完成所有表信息的添加后，保存CSV文件的更改。</li> <li>5. 将编辑并保存好的CSV文件上传至MgC。</li> </ol> |
| 目的端配置 | 目的端连接 | <p>选择创建的<b>目的端DLI 通用队列连接</b>。</p> <p><b>注意</b><br/>切勿选择DLI SQL队列。</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |

| 区域                                       | 参数                                          | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                                         |        |   |                                        |        |   |                             |                               |   |                                    |                                  |   |
|------------------------------------------|---------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------|------|---|---------------------------|--|---|-----------------------------------------|------------------------------------------|---|------------------------------------------|---------------------------------------------|---|-----------------------------|------|---|---------------------------|--|---|-----------------------------------------|------------------------------------------|---|------------------------------------------|---------------------------------------------|---|-----------------------------------------|--------|---|----------------------------------------|--------|---|-----------------------------|-------------------------------|---|------------------------------------|----------------------------------|---|
|                                          | 自定义参数（可选）                                   | <p>根据实际需求进行配置，支持的自定义参数详情请参考<a href="#">conf参数说明</a>和<a href="#">自定义参数说明</a>。</p> <ul style="list-style-type: none"> <li>如果是<b>公网迁移</b>，需要填写以下4个参数：           <div data-bbox="858 421 1369 629" style="border: 1px solid #ccc; padding: 5px; margin: 10px 0;"> <p>自定义参数<br/>参数详情请参考文档中“表3 conf参数说明”部分。 <a href="#">查看文档</a></p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 60%;">spark.dli.metaAccess.enable</td> <td>true</td> <td style="text-align: right;">✕</td> </tr> <tr> <td>spark.dli.job.agency.name</td> <td></td> <td style="text-align: right;">✕</td> </tr> <tr> <td>mgc.mc2dli.data.migration.dli.file.path</td> <td>obs://mgc-test/data/migration-dli-spark-</td> <td style="text-align: right;">✕</td> </tr> <tr> <td>mgc.mc2dli.data.migration.dli.spark.jars</td> <td>["obs://mgc-test/data/datasource.jar","oi"]</td> <td style="text-align: right;">✕</td> </tr> </table> <p style="text-align: center; color: #00a0e3; font-size: small;">添加参数</p> </div> <ul style="list-style-type: none"> <li>- spark.dli.metaAccess.enable：填写true。</li> <li>- spark.dli.job.agency.name：填写在完成准备工作时配置的<b>DLI委托权限名称</b>。</li> <li>- mgc.mc2dli.data.migration.dli.file.path：填写在完成准备工作时<b>存放migration-dli-spark-1.0.0.jar包的OBS路径</b>。例如：<br/>obs://mgc-test/data/migration-dli-spark-1.0.0.jar</li> <li>- mgc.mc2dli.data.migration.dli.spark.jars：填写在完成准备工作时<b>存放fastjson-1.2.54.jar和datasource.jar包的OBS路径</b>，以数组形式传入，路径带双引号，以英文逗号隔开。例如：["obs://mgc-test/data/datasource.jar","obs://mgc-test/data/fastjson-1.2.54.jar"]</li> </ul> <li>如果是<b>专线迁移</b>，需要填写以下8个参数：           <div data-bbox="858 1238 1369 1597" style="border: 1px solid #ccc; padding: 5px; margin: 10px 0;"> <p>自定义参数<br/>参数详情请参考文档中“表3 conf参数说明”部分。 <a href="#">查看文档</a></p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 60%;">spark.dli.metaAccess.enable</td> <td>true</td> <td style="text-align: right;">✕</td> </tr> <tr> <td>spark.dli.job.agency.name</td> <td></td> <td style="text-align: right;">✕</td> </tr> <tr> <td>mgc.mc2dli.data.migration.dli.file.path</td> <td>obs://mgc-test/data/migration-dli-spark-</td> <td style="text-align: right;">✕</td> </tr> <tr> <td>mgc.mc2dli.data.migration.dli.spark.jars</td> <td>["obs://mgc-test/data/datasource.jar","oi"]</td> <td style="text-align: right;">✕</td> </tr> <tr> <td>spark.sql.catalog.mc_catalog.tableWrite</td> <td>tunnel</td> <td style="text-align: right;">✕</td> </tr> <tr> <td>spark.sql.catalog.mc_catalog.tableReac</td> <td>tunnel</td> <td style="text-align: right;">✕</td> </tr> <tr> <td>spark.hadoop.odps.end.point</td> <td>http://service.ci. maxcompute</td> <td style="text-align: right;">✕</td> </tr> <tr> <td>spark.hadoop.odps.tunnel.end.point</td> <td>http://dt.ci. j.maxcompute.allyt</td> <td style="text-align: right;">✕</td> </tr> </table> <p style="text-align: center; color: #00a0e3; font-size: small;">添加参数</p> </div> <ul style="list-style-type: none"> <li>- spark.dli.metaAccess.enable：填写true。</li> <li>- spark.dli.job.agency.name：填写在完成准备工作时配置的<b>DLI委托权限名称</b>。</li> <li>- mgc.mc2dli.data.migration.dli.file.path：填写在完成准备工作时<b>存放migration-dli-spark-1.0.0.jar包的OBS路径</b>。例如：<br/>obs://mgc-test/data/migration-dli-spark-1.0.0.jar</li> <li>- mgc.mc2dli.data.migration.dli.spark.jars：填写在完成准备工作时<b>存放</b></li> </ul> </li> </li></ul> | spark.dli.metaAccess.enable | true | ✕ | spark.dli.job.agency.name |  | ✕ | mgc.mc2dli.data.migration.dli.file.path | obs://mgc-test/data/migration-dli-spark- | ✕ | mgc.mc2dli.data.migration.dli.spark.jars | ["obs://mgc-test/data/datasource.jar","oi"] | ✕ | spark.dli.metaAccess.enable | true | ✕ | spark.dli.job.agency.name |  | ✕ | mgc.mc2dli.data.migration.dli.file.path | obs://mgc-test/data/migration-dli-spark- | ✕ | mgc.mc2dli.data.migration.dli.spark.jars | ["obs://mgc-test/data/datasource.jar","oi"] | ✕ | spark.sql.catalog.mc_catalog.tableWrite | tunnel | ✕ | spark.sql.catalog.mc_catalog.tableReac | tunnel | ✕ | spark.hadoop.odps.end.point | http://service.ci. maxcompute | ✕ | spark.hadoop.odps.tunnel.end.point | http://dt.ci. j.maxcompute.allyt | ✕ |
| spark.dli.metaAccess.enable              | true                                        | ✕                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                                         |        |   |                                        |        |   |                             |                               |   |                                    |                                  |   |
| spark.dli.job.agency.name                |                                             | ✕                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                                         |        |   |                                        |        |   |                             |                               |   |                                    |                                  |   |
| mgc.mc2dli.data.migration.dli.file.path  | obs://mgc-test/data/migration-dli-spark-    | ✕                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                                         |        |   |                                        |        |   |                             |                               |   |                                    |                                  |   |
| mgc.mc2dli.data.migration.dli.spark.jars | ["obs://mgc-test/data/datasource.jar","oi"] | ✕                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                                         |        |   |                                        |        |   |                             |                               |   |                                    |                                  |   |
| spark.dli.metaAccess.enable              | true                                        | ✕                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                                         |        |   |                                        |        |   |                             |                               |   |                                    |                                  |   |
| spark.dli.job.agency.name                |                                             | ✕                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                                         |        |   |                                        |        |   |                             |                               |   |                                    |                                  |   |
| mgc.mc2dli.data.migration.dli.file.path  | obs://mgc-test/data/migration-dli-spark-    | ✕                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                                         |        |   |                                        |        |   |                             |                               |   |                                    |                                  |   |
| mgc.mc2dli.data.migration.dli.spark.jars | ["obs://mgc-test/data/datasource.jar","oi"] | ✕                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                                         |        |   |                                        |        |   |                             |                               |   |                                    |                                  |   |
| spark.sql.catalog.mc_catalog.tableWrite  | tunnel                                      | ✕                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                                         |        |   |                                        |        |   |                             |                               |   |                                    |                                  |   |
| spark.sql.catalog.mc_catalog.tableReac   | tunnel                                      | ✕                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                                         |        |   |                                        |        |   |                             |                               |   |                                    |                                  |   |
| spark.hadoop.odps.end.point              | http://service.ci. maxcompute               | ✕                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                                         |        |   |                                        |        |   |                             |                               |   |                                    |                                  |   |
| spark.hadoop.odps.tunnel.end.point       | http://dt.ci. j.maxcompute.allyt            | ✕                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                             |      |   |                           |  |   |                                         |                                          |   |                                          |                                             |   |                                         |        |   |                                        |        |   |                             |                               |   |                                    |                                  |   |

| 区域   | 参数      | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
|------|---------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|      |         | <p><b>fastjson-1.2.54.jar和datasource.jar包的OBS路径</b>，以数组形式传入，路径带双引号，以英文逗号隔开。例如：["obs://mgc-test/data/datasource.jar","obs://mgc-test/data/fastjson-1.2.54.jar"]</p> <ul style="list-style-type: none"> <li>- spark.sql.catalog.mc_catalog.tableWriteProvider: 填写tunnel。</li> <li>- spark.sql.catalog.mc_catalog.tableReadProvider: 填写tunnel。</li> <li>- spark.hadoop.odps.end.point: 填写源端MaxCompute服务所在地域的<b>VPC网络Endpoint</b>信息。各地域<b>VPC网络Endpoint</b>信息请参考<b>Endpoint对照表</b>。例如：源端所在地域为“香港”，则填写“http://service.cn-hongkong.maxcompute.aliyun-inc.com/api”</li> <li>- spark.hadoop.odps.tunnel.end.point: 填写源端MaxCompute服务所在地域的<b>VPC网络Tunnel Endpoint</b>信息。各地域<b>VPC网络Tunnel Endpoint</b>信息请参考<b>Endpoint对照表</b>。例如：源端所在地域为“香港”，则填写“http://dt.cn-hongkong.maxcompute.aliyun-inc.com”</li> </ul> |
| 迁移选项 | 大表规则    | 设置大表条件，当符合大表的条件时，将自动按分区拆分迁移子任务。建议保持默认，也可根据实际需求修改。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|      | 小表规则    | 设置小表条件，当符合小表的条件时，自动将多个小表合并到一个迁移子任务中，提高迁移效率。建议保持默认，也可根据实际需求修改。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|      | 并发连接数   | 设置迁移并发数。默认为3，1≤取值范围≤10。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|      | 单文件SQL数 | 每批提交的文件中所包含的SQL条数。默认为3，1≤推荐取值范围≤50。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |

**步骤6** 配置完成后，选择任务的运行方式。

**须知**

- 创建的迁移任务可重复运行，每运行一次，即会生成一条任务实例。
- 单击任务名称，可以修改任务配置。
- **立即运行**：如果需要立即运行迁移任务，可以勾选“**保存后立即运行**”，然后单击“**保存**”按钮，会在**任务管理**列表中生成一条增量数据迁移任务并自动开始执行迁移。



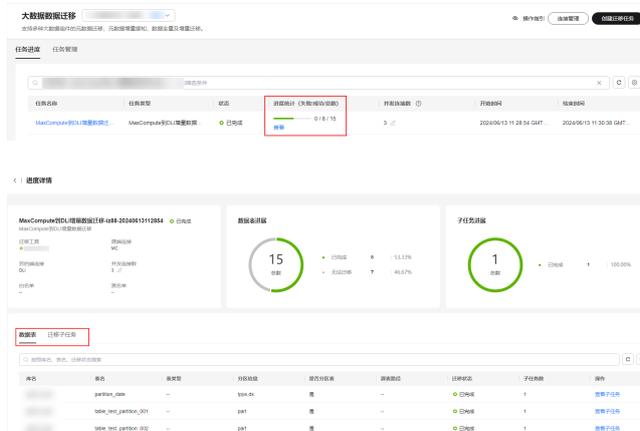
- **手动运行**：如果需要后续手动启动迁移任务，则只需单击“保存”按钮，在任务管理列表中生成一条增量数据迁移任务。在需要启动迁移时，单击操作列的“运行”，手动启动迁移。



**步骤7** 迁移任务运行后，单击操作列的“查看实例”，进入任务进度页签，可以查看该任务的迁移进度以及产生的所有运行实例。



单击进度统计列的“查看”，进入进度详情页面。可以查看迁移的数据表清单、状态并导出任务结果。



**步骤8** （可选）数据迁移完成后，可以使用MgC提供的大数据校验功能，对源端和目的端数据进行一致性校验。详细的操作步骤请参考[MaxCompute迁移至DLI数据校验](#)。

----结束

## 1.11.7 查看并导出执行结果

每运行一次迁移任务，即会生成一条任务实例。在任务实例的进度详情页面，可以查看并导出任务的执行结果。

## 须知

导出的结果文件可以在生成的15天内下载，15天后自动删除。

## 前提条件

已创建迁移任务并运行。

## 操作步骤

**步骤1** 在任务进度列表，单击进度统计列的“查看”，进入**进度详情**页面。



**步骤2** 单击列表上方的“导出”按钮，系统开始导出任务执行结果。通过列表上方的搜索框可以对任务执行结果进行过滤，如有过滤，则只导出过滤后的数据。

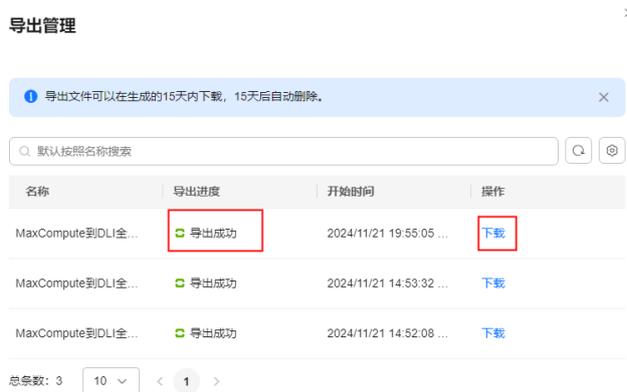
图 1-57 导出执行结果



**步骤3** 单击“导出管理”按钮，右侧弹出**导出管理**窗口。

**步骤4** 当导出进度为导出成功后，单击操作列的“下载”，可以将导出结果下载到本地。

图 1-58 下载导出结果



---结束

### 1.11.8 自定义参数说明

在目的端配置中，支持配置的自定义参数及说明请参见[表1-71](#)和[conf参数说明](#)。

表 1-71 支持的自定义参数说明

| 参数名称                                            | 取值范围 | 说明                                                                                                                                                                | 是否必填      |
|-------------------------------------------------|------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| spark.dli.metaAccess.enable                     | -    | 填写true。                                                                                                                                                           | 是         |
| spark.dli.job.agency.name                       | -    | 填写 <b>DLI委托权限名称</b> 。                                                                                                                                             |           |
| mgc.mc2dli.data.migration.dli.file.path         | -    | 填写 <b>存放migration-dli-spark-1.0.0.jar包的OBS路径</b> 。例如：obs://mgc-test/data/migration-dli-spark-1.0.0.jar                                                            |           |
| mgc.mc2dli.data.migration.dli.spark.jars        | -    | 填写 <b>存放fastjson-1.2.54.jar和datasource.jar包的OBS路径</b> ，以数组形式传入，路径带双引号，以英文逗号隔开。例如：["obs://mgc-test/data/datasource.jar","obs://mgc-test/data/fastjson-1.2.54.jar"] |           |
| spark.sql.catalog.mc_catalog.tableWriteProvider | -    | 填写tunnel                                                                                                                                                          | 专线迁移场景为必填 |
| spark.sql.catalog.mc_catalog.tableReadProvider  | -    | 填写tunnel                                                                                                                                                          |           |

| 参数名称                                        | 取值范围                                                                      | 说明                                                                                                                                                                                                                                                                                                                                                                                                                                 | 是否必填 |
|---------------------------------------------|---------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| spark.hadoop.odps.end.point                 |                                                                           | 填写源端MaxCompute服务所在地域的 <b>VPC网络Endpoint</b> 信息。各地域 <b>VPC网络Endpoint</b> 信息请参考 <a href="#">Endpoint对照表</a> 。例如：源端所在地域为“香港”，则填写“http://service.cn-hongkong.maxcompute.aliyun-inc.com/api”                                                                                                                                                                                                                                             |      |
| spark.hadoop.odps.tunnel.end.point          |                                                                           | 填写源端MaxCompute服务所在地域的 <b>VPC网络Tunnel Endpoint</b> 信息。各地域 <b>VPC网络Tunnel Endpoint</b> 信息请参考 <a href="#">Endpoint对照表</a> 。例如：源端所在地域为“香港”，则填写“http://dt.cn-hongkong.maxcompute.aliyun-inc.com”                                                                                                                                                                                                                                        |      |
| spark.hadoop.odps.tunnel.quota.name         | -                                                                         | 填写访问MaxCompute使用的Quota名称。                                                                                                                                                                                                                                                                                                                                                                                                          | 否    |
| mgc.mc2dli.data.migration.mc.external.table | <ul style="list-style-type: none"> <li>● false</li> <li>● true</li> </ul> | 设置是否迁移外表。 <ul style="list-style-type: none"> <li>● false不迁移外表。</li> <li>● true迁移外表。</li> </ul>                                                                                                                                                                                                                                                                                                                                     | 否    |
| sc_type                                     | A、B、C                                                                     | 计算资源类型，目前指定参数A, B, C。如果不指定，则按最小类型创建。 <ul style="list-style-type: none"> <li>● A: (8核32G内存, driverCores: 2, executorCores: 1, driverMemory: 7G, executorMemory: 4G, numExecutor: 6)</li> <li>● B: (16核64G内存, driverCores: 2, executorCores: 2, driverMemory: 7G, executorMemory: 8G, numExecutor: 7)</li> <li>● C: (32核128G内存, driverCores: 4, executorCores: 2, driverMemory: 15G, executorMemory: 8G, numExecutor: 14)</li> </ul> | 否    |

| 参数名称                             | 取值范围                                                                            | 说明                                                                             | 是否必填 |
|----------------------------------|---------------------------------------------------------------------------------|--------------------------------------------------------------------------------|------|
| executorCores                    | 1- 4                                                                            | Spark应用每个Executor的CPU核数。该配置项会替换“sc_type”中对应的默认参数。                              |      |
| numExecutors                     | 1-100                                                                           | Spark应用Executor的个数。该配置项会替换“sc_type”中对应的默认参数。                                   |      |
| executorMemory                   | 1-16 G                                                                          | Spark应用的Executor内存，参数配置例如：2G, 2048M。该配置项会替换“sc_type”中对应的默认参数，使用时必须带单位，否则会启动失败。 |      |
| driverCores                      | 1-4                                                                             | Spark应用Driver的CPU核数。该配置项会替换“sc_type”中对应的默认参数。                                  |      |
| driverMemory                     | 1-16 G                                                                          | Spark应用的Driver内存，参数配置例如：2G, 2048M。该配置项会替换“sc_type”中对应的默认参数，使用时必须带单位，否则会启动失败。   |      |
| hive.exec.dynamic.partition.mode | <ul style="list-style-type: none"> <li>• strict</li> <li>• nonstrict</li> </ul> | 选择使用“自定义过滤”迁移策略迁移动态分区数据时，请添加该参数并配置参数值为 <b>nonstrict</b> 。                      | 否    |

## 1.12 大数据校验

### 1.12.1 大数据校验概述

迁移中心 MgC为多种大数据计算/存储引擎，如Hive、Hbase、Doris和MaxCompute等，提供了一致性校验方法。通过一致性校验，您可以更加放心地进行大数据迁移操作，确保数据的准确性和可靠性，从而避免了数据丢失或不一致的风险。

#### 注意事项

- 创建的源端统计任务和目的端统计任务的统计方式必须保持一致。
- HBase安全集群任务与非安全集群任务不能同时执行。HBase安全集群需要加载认证信息，非安全集群连接时需要清除安全集群的认证信息，因此不能同时执行。
- 当源端服务（包含但不限于Lindorm、HBase）因欠费被锁定时，虽然可以正常创建数据连接和校验任务，但由于服务锁定限制了数据访问和操作权限，最终会导致校验任务无法正常执行。在进行数据校验前，请确保源端服务未因欠费被锁定，以避免权限问题。如果服务被锁定，请及时补缴费用以解除锁定。确认服务恢复正常后，重新尝试执行数据校验任务。

- 在进行Hive 2.x与Hive 3.x版本间数据校验时，由于Hive 2.x版本在查询固定长度字符串类型char(N)数据时，如果实际数据长度不满足N长度，Hive会以空格补齐至N长度。而Hive 3.x版本在查询时不会进行这种补齐操作。这可能导致在不同版本间比较结果时出现差异。为了避免因版本差异导致的数据不一致问题，建议使用Beeline作为客户端工具进行作业对数。
- 在完成MRS到MRS集群数据迁移后，如果源端和目的端都使用YARN模式进行校验，需要分开执行校验任务，确保一个任务执行完成后才能执行另一个任务。
- 在进行对数任务时，不支持使用 MRS 3.3.0 及以上版本集群节点的ECS作为执行机，会导致对数任务执行失败。

## 约束与限制

- 创建EMR Delta Lake到MRS Delta Lake数据校验任务存在以下约束与限制：
  - 源端EMR环境的Spark版本为3.3.1时，支持Delta Lake（有元数据）和Delta Lake（无元数据）的数据校验。
  - 源端EMR环境的Spark版本为2.4.8时，仅支持Delta Lake（有元数据）的数据校验。
- Hbase校验不支持纯冷数据表。
- 校验任务不支持跨天执行，执行过程中如果跨0点将导致校验结果异常，请合理规划任务以避免跨天运行。
- 在进行ClickHouse数据校验时，如果源端阿里云的ClickHouse版本为21.8.15.7，目的端华为云的ClickHouse版本为23.3.2.37，由于这两个版本在处理IPv4和IPv6字段类型以及相关函数计算结果上存在差异，因此无法进行字段校验。
- 在执行Hive数据的日级别增量、小时级增量以及指定日期的校验时，如果分区字段为date类型，且分区值的格式不是标准的YYYY-MM-DD时，系统目前不支持对这种场景进行数据校验。
- 不支持两端均为HBase 2.x不同版本的安全集群数据校验，因为不同版本间存在版本兼容性限制、安全认证机制差异、协议和接口不一致以及特性支持和配置不同等问题，都会影响校验结果。

## 校验方式说明

- **全量校验**：全量对比所有数据，适用于需要对比历史数据的场景。
- **日级增量校验**：依靠元数据中的创建或更新时间，获取数据变化（增量）分区，再进行校验。支持单日或连续多日增量数据的统计，是最常用的校验方式。
- **小时级增量校验**：依靠元数据中的创建或更新时间，获取数据变化（增量）分区，再进行校验。在24小时内自动进行多次校验，可以持续关注数据一致性变化情况。第二日0点自动停止任务。
- **指定日期校验**：仅适用于分区为年月日格式的分区表，支持周期性指定一个或多个连续日期分区进行校验。分区格式不为年月日的表不会进行校验。
- **指定时间校验**：用户可以选择一个时间段，校验该时间段内的数据。只能选择当前时间之前的时间段进行校验。

## 支持的源端、目的端组件

| 源端组件                                                                                                                                                                         | 目的端组件                                                                                                                                                                                                                                          |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> <li>• Hive</li> <li>• Hbase</li> <li>• Doris</li> <li>• MaxCompute</li> <li>• ClickHouse</li> <li>• Delta Lake</li> <li>• Hudi</li> </ul> | <ul style="list-style-type: none"> <li>• Hive</li> <li>• DLI</li> <li>• MRS(Doris)</li> <li>• MRS(Hbase)</li> <li>• MRS(ClickHouse)</li> <li>• CloudTable(ClickHouse)</li> <li>• CloudTable(HBase)</li> <li>• Delta</li> <li>• Hudi</li> </ul> |

## 各组件支持的校验方式

| 组件                     | 支持的校验方式                                                                                                         |
|------------------------|-----------------------------------------------------------------------------------------------------------------|
| Hive                   | <ul style="list-style-type: none"> <li>• 全量校验</li> </ul>                                                        |
| DLI                    | <ul style="list-style-type: none"> <li>• 日级增量校验</li> <li>• 小时级增量校验</li> <li>• 指定日期校验</li> </ul>                 |
| MaxCompute             | <ul style="list-style-type: none"> <li>• 全量校验</li> <li>• 日级增量校验</li> <li>• 小时级增量校验</li> <li>• 指定日期校验</li> </ul> |
| Doris                  | <ul style="list-style-type: none"> <li>• 全量校验</li> <li>• 日级增量校验</li> <li>• 小时级增量校验</li> </ul>                   |
| Hbase                  | <ul style="list-style-type: none"> <li>• 全量校验</li> <li>• 指定时间校验</li> </ul>                                      |
| ClickHouse             | 全量校验                                                                                                            |
| 阿里云云数据库 ClickHouse     | 全量校验                                                                                                            |
| CloudTable(HBase)      | <ul style="list-style-type: none"> <li>• 全量校验</li> <li>• 指定时间校验</li> </ul>                                      |
| CloudTable(ClickHouse) | 全量校验                                                                                                            |

| 组件    | 支持的校验方式                                                                                                    |
|-------|------------------------------------------------------------------------------------------------------------|
| Delta | <ul style="list-style-type: none"><li>● 全量校验</li><li>● 日级增量校验</li><li>● 小时级增量校验</li><li>● 指定日期校验</li></ul> |
| Hudi  | <ul style="list-style-type: none"><li>● 全量校验</li><li>● 日级增量校验</li><li>● 小时级增量校验</li><li>● 指定日期校验</li></ul> |

## 1.12.2 准备工作

在使用大数据校验功能前，请务必先完成以下必要的准备工作。

### 华为账号准备

使用MgC之前，您需要拥有一个华为账号或一个可用于访问MgC的IAM用户，并获取账号/IAM用户的访问密钥（AK/SK）。获取方法请参考[准备工作](#)。

#### 说明

在创建元数据连接时，如果要使用“采集资源使用信息”功能和查看资源概览，需要为IAM账号的AK/SK添加MRS服务和DLI服务的只读权限（ReadOnly）。

### 创建大数据迁移项目

通过创建独立的项目，可以将不同的操作场景进行区分和隔离，从而实现更好的组织管理和资源分配。创建方法请参考[项目管理](#)。

### 准备 MgC Agent（原 Edge）

在启用大数据校验功能前，需要先在源端环境中安装用于配合大数据校验的MgC Agent。大数据校验场景仅支持Linux版本的MgC Agent，安装方法请参考[下载并安装Linux版本MgC Agent](#)。安装成功后，将MgC Agent与迁移中心进行连接，连接方法请参考[连接迁移中心](#)。

---

#### 注意

- MgC Agent与迁移中心连接时，请将输入的华为云账号AK/SK保存为目的端凭证。
  - MgC Agent不具备自动重启功能，严禁在任务执行期间重启MgC Agent，以避免任务失败。
-

## 选择并启用 MgC Agent (原 Edge)

- 步骤1** MgC Agent与迁移中心连接成功后，登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择创建的[大数据迁移项目](#)。
- 步骤2** 在左侧导航栏选择“[实施>大数据校验](#)”，进入[大数据校验](#)页面。
- 步骤3** 单击“[选择并启用迁移中心Agent](#)”按钮，弹出[选择迁移中心Agent](#)窗口。
- 步骤4** 在下拉列表中，选择与迁移中心连接成功的MgC Agent。
- 如迁移工具状态为“[在线](#)”、“[已启用](#)”，代表该工具已被用于对数相关操作，单击“[确定](#)”按钮，继续使用该工具进行后续操作。
  - 如迁移工具状态为“[在线](#)”、“[未启用](#)”，代表该工具尚未被用于对数，单击“[启用](#)”按钮，启用该工具进行后续操作。
  - 如迁移工具状态为“[离线](#)”、“[已删除](#)”，代表该工具已被删除，无法进行对数操作；如果该工具在删除前进行过对数操作，可以单击“[查看](#)”按钮，查看已有数据。
  - 如迁移工具状态为“[离线](#)”、“[未启用](#)”，代表该工具无法进行对数相关操作，单击“[查看](#)”按钮，界面上的所有操作按钮会被禁用。



**注意**

在进行大数据校验前，请确保MgC Agent处于在线并启用状态。

---

----结束

## 1.12.3 创建源端连接

### 1.12.3.1 创建源端云服务连接

对大数据云服务数据进行一致性校验时，需要分别与源端和目的端云服务建立连接，用于数据查询和计算。

目前支持创建源端连接的大数据云服务包括：

- [MaxCompute](#)
- [阿里云云数据库ClickHouse](#)

### 操作步骤

- 步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择创建的[大数据迁移项目](#)。
- 步骤2** 在左侧导航栏选择“[实施>大数据校验](#)”，进入[大数据校验](#)页面。
- 步骤3** 首次使用大数据校验功能时，需要先启用MgC Agent。单击“[选择并启用迁移中心Agent](#)”按钮，在弹窗的下拉列表中，选择完成准备工作时与迁移中心连接成功的[MgC Agent](#)。

**注意**

在进行大数据校验时，请确保MgC Agent一直处于在线并启用状态。

**步骤4** 在功能入口区域，单击“**迁移准备**”，进入**迁移准备**页面。

**步骤5** 选择“**连接管理**”，单击“**创建连接**”，右侧弹出**新增连接**窗口。

图 1-59 创建连接



**步骤6** 根据实际情况选择大数据云服务，单击“**下一步**”，进入**详细配置**页面。

**步骤7** 根据对应的参数配置说明，配置云服务参数。

- [MaxCompute连接参数配置说明](#)
- [阿里云云数据库ClickHouse连接参数配置说明](#)

表 1-72 MaxCompute 连接参数配置说明

| 参数        | 配置说明                                                                         |
|-----------|------------------------------------------------------------------------------|
| 所属源       | 选择源端。                                                                        |
| 连接名称      | 默认创建名称为“MaxCompute-4位随机数（字母和数字）”，也可以自定义名称。                                   |
| 迁移中心Agent | 选择在源端环境中安装的 <b>MgC Agent（原Edge）</b> 。                                        |
| 阿里云凭证     | 选择源端阿里云账号的凭证。凭证的添加方式请参考 <a href="#">添加资源凭证</a> ，类型选择“大数据-MaxCompute”。        |
| 项目名称      | 输入需要采集的MaxCompute项目名称。项目名称可在MaxCompute侧控制台获取。                                |
| Endpoint  | 输入MaxCompute服务所在地域的Endpoint信息。各地域Endpoint信息请参考 <a href="#">Endpoint对照表</a> 。 |

表 1-73 阿里云云数据库 ClickHouse 连接参数配置说明

| 参数        | 配置说明                                              |
|-----------|---------------------------------------------------|
| 所属源       | 选择源端。                                             |
| 连接名称      | 默认创建名称为“阿里云云数据库ClickHouse-4位随机数（字母和数字）”，也可以自定义名称。 |
| 迁移中心Agent | 选择在源端环境中安装的 <b>MgC Agent（原Edge）</b> 。             |

| 参数           | 配置说明                                                                                    |
|--------------|-----------------------------------------------------------------------------------------|
| ClickHouse凭证 | 选择在MgC Agent添加的源端阿里云数据库ClickHouse凭证。凭证的添加方法请参考 <a href="#">添加资源凭证</a> 中的大数据-ClickHouse。 |
| 数据库链接        | 填写连接源端ClickHouse集群的外网地址。可以在 <a href="#">集群信息</a> 页面，查看外网地址。                             |

**步骤8** 参数配置完成后，单击“**测试连接**”，系统会校验配置信息并与云服务建立连接，测试通过代表与云服务连接成功。

**步骤9** 测试连接通过后，单击“**确认**”按钮，创建云服务连接完成。

**步骤10** 在**连接管理**页面，可以查看创建的连接列表以及基本信息。单击操作列的“**修改连接**”，可以修改连接参数配置。

---结束

### 1.12.3.2 创建源端组件连接

对大数据组件数据进行一致性校验时，需要与大数据组件建立连接，用于数据查询。

目前支持创建源端连接的大数据组件包括：

- Doris
- HBase
- ClickHouse
- Hive Metastore
- Delta Lake（有元数据）
- Delta Lake（无元数据）
- Hudi（有元数据）
- Hudi（无元数据）

### 操作步骤

**步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择创建的[大数据迁移项目](#)。

**步骤2** 在左侧导航栏选择“**实施>大数据校验**”，进入**大数据校验**页面。

**步骤3** 首次使用大数据校验功能时，需要先启用MgC Agent。单击“**选择并启用迁移中心Agent**”按钮，在弹窗的下拉列表中，选择完成准备工作时与迁移中心连接成功的[MgC Agent](#)。

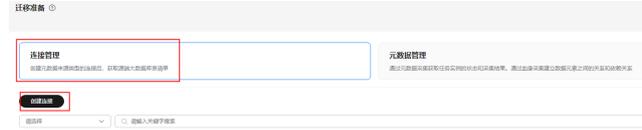
#### 注意

在进行大数据校验时，请确保MgC Agent一直处于在线并启用状态。

**步骤4** 在功能入口区域，单击“迁移准备”，进入迁移准备页面。

**步骤5** 选择“连接管理”，单击“创建连接”，右侧弹出新增连接窗口。

图 1-60 创建连接



**步骤6** 根据实际情况选择大数据组件，单击“下一步”，进入详细配置页面。

**步骤7** 根据对应的参数配置说明，配置执行机参数。

- [Doris连接参数配置说明](#)
- [HBase连接参数配置说明](#)
- [ClickHouse组件参数配置说明](#)
- [Hive Metastore连接参数配置说明](#)
- [Delta Lake（有元数据）连接参数配置说明](#)
- [Delta Lake（无元数据）连接参数配置说明](#)
- [Hudi（有元数据）连接参数配置说明](#)
- [Hudi（无元数据）连接参数配置说明](#)

表 1-74 Doris 连接参数配置说明

| 参数        | 配置说明                                                                         |
|-----------|------------------------------------------------------------------------------|
| 所属源       | 选择源端。                                                                        |
| 连接名称      | 默认创建名称为“Doris-4位随机数（字母和数字）”，也可以自定义名称。                                        |
| 迁移中心Agent | 选择在源端环境中安装的 <b>MgC Agent（原Edge）</b> 。                                        |
| Doris凭证   | 选择在MgC Agent添加的源端Doris凭证。凭证的添加方式请参考 <a href="#">添加资源凭证</a> ，类型选择“大数据-Doris”。 |
| 数据库地址     | 填写连接到源端Doris集群部署的服务器IP地址。                                                    |
| 数据库端口     | 填写连接到源端Doris集群部署的服务器的端口。默认为3306。                                             |
| 数据库名      | 填写源端Doris数据库名称。                                                              |

| 参数       | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 采集资源使用信息 | <p>可选参数。开启后，通过该连接创建的统计任务将收集大数据服务的资源使用信息，这些信息将用于MgC上的报告生成，以便于进行性能优化。</p> <p><b>须知</b><br/>使用该功能，需要MgC Agent添加的华为云账号AK/SK拥有MRS服务和DLI服务的只读权限（ReadOnly）。</p> <ul style="list-style-type: none"> <li>● 如果所选凭证与登录迁移中心 MgC的账号凭证相同，请勾选“<b>所选凭证与登录迁移中心凭证相同</b>”，系统将根据所选的区域，自动查询账号在该区域下的项目。 <ul style="list-style-type: none"> <li>- 区域：选择需要校验的数据所在区域/项目。</li> <li>- 项目：选择需要校验的数据所在项目。</li> <li>- 集群ID：输入需要校验的数据所在集群ID。</li> </ul> </li> <li>● 如果所选凭证与登录迁移中心 MgC的账号凭证不同，请手动输入如下参数信息： <ul style="list-style-type: none"> <li>- 区域ID：输入需要校验的数据所在区域ID。例如，区域为“华南-广州”，则输入“cn-south-1”</li> <li>- 项目ID：输入与区域对应的项目ID。</li> <li>- 集群ID：输入需要校验的数据所在集群ID。</li> </ul> </li> </ul> <p><b>说明</b></p> <ul style="list-style-type: none"> <li>● 区域ID和项目ID可以在“我的凭证 &gt; <a href="#">API凭证</a>”查看。</li> <li>● 集群ID的获取方法请参考获取<a href="#">MRS集群ID</a>。</li> </ul> |

表 1-75 HBase 连接参数配置说明

| 参数             | 配置说明                                                                         |
|----------------|------------------------------------------------------------------------------|
| 所属源            | 选择源端。                                                                        |
| 连接名称           | 默认创建名称为“HBase-4位随机数（字母和数字）”，也可以自定义名称。                                        |
| 迁移中心Agent      | 选择在源端环境中安装的 <b>MgC Agent（原Edge）</b> 。                                        |
| HBase凭证        | 选择在MgC Agent添加的源端HBase凭证。凭证的添加方式请参考 <a href="#">添加资源凭证</a> ，类型选择“大数据-HBase”。 |
| 安全集群           | 选择是否部署在安全集群下。                                                                |
| Zookeeper节点IP  | 填写连接到源端Zookeeper的IP地址。通常是部署Zookeeper的服务器公网IP地址或内网IP地址。                       |
| Zookeepere连接端口 | 填写连接到源端Zookeepere的端口。                                                        |
| HBase 版本       | 选择源端HBase的版本。                                                                |

表 1-76 ClickHouse 连接参数配置说明

| 参数                   | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|----------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 所属源                  | 选择源端。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| 连接名称                 | 默认创建名称为“ClickHouse-4位随机数（字母和数字）”，也可以自定义名称。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| 迁移中心Agent            | 选择在源端环境中安装的 <b>MgC Agent（原Edge）</b> 。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| ClickHouse凭证（可选）     | 选择在MgC Agent添加的源端ClickHouse凭证。凭证的添加方法请参考 <a href="#">添加资源凭证</a> 中的大数据-ClickHouse。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| 安全集群                 | 选择是否部署在安全集群下。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| ClickHouse Server IP | 填写连接到源端ClickHouse服务器的IP地址。通常是运行ClickHouse的主机IP地址。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| HTTP端口               | 非安全集群下，与源端ClickHouse 服务器通信的 HTTP 端口。<br>可以登录源端集群的 <b>FusionInsight Manager</b> ，选择“集群 > 服务 > ClickHouse > 配置 > 全部配置”，搜索“http_port”参数获取。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| HTTP SSL/TLS端口       | 安全集群下，与源端ClickHouse 服务器通信的HTTPS 端口。<br>可以登录源端集群的 <b>FusionInsight Manager</b> ，选择“集群 > 服务 > ClickHouse > 配置 > 全部配置”，搜索“https_port”参数获取。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| 采集资源使用信息             | <p>可选参数。开启后，通过该连接创建的统计任务将收集大数据服务的资源使用信息，这些信息将用于MgC上的报告生成，以便于进行性能优化。</p> <p><b>须知</b></p> <p>使用该功能，需要MgC Agent添加的华为云账号AK/SK拥有MRS服务和DLI服务的只读权限（ReadOnly）。</p> <ul style="list-style-type: none"> <li>如果所选凭证与登录迁移中心 MgC的账号凭证相同，请勾选“<b>所选凭证与登录迁移中心凭证相同</b>”，系统将根据所选的区域，自动查询账号在该区域下的项目。 <ul style="list-style-type: none"> <li>区域：选择需要校验的数据所在区域/项目。</li> <li>项目：选择需要校验的数据所在项目。</li> <li>集群ID：输入需要校验的数据所在集群ID。</li> </ul> </li> <li>如果所选凭证与登录迁移中心 MgC的账号凭证不同，请手动输入如下参数信息： <ul style="list-style-type: none"> <li>区域ID：输入需要校验的数据所在区域ID。例如，区域为“华南-广州”，则输入“cn-south-1”</li> <li>项目ID：输入与区域对应的项目ID。</li> <li>集群ID：输入需要校验的数据所在集群ID。</li> </ul> </li> </ul> <p><b>说明</b></p> <ul style="list-style-type: none"> <li>区域ID和项目ID可以在“我的凭证 &gt; <b>API凭证</b>”查看。</li> <li>集群ID的获取方法请参考获取<b>MRS集群ID</b>。</li> </ul> |

表 1-77 Hive Metastore 连接参数配置说明

| 参数                   | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 所属源                  | 选择源端。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| 连接名称                 | 默认创建名称为“Hive-Metastore-4位随机数（字母和数字）”，也可以自定义名称。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| 启用安全连接               | 选择是否启用安全连接。 <ul style="list-style-type: none"> <li>• MetaStore在非安全集群下，可以不启用安全连接。</li> <li>• MetaStore在安全集群下，需要启用安全连接并选择凭证。凭证的获取以及添加方法请参考<a href="#">添加资源凭证</a>中的大数据-Hive Metastore。</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| Hive版本               | 选择源端Hive的实际版本。<br><b>注意</b><br>如果源端Hive为2.1.1版本，此处需要选择1.x版本。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| MetaStore节点IP        | 填写连接MetaStore的IP地址。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| MetaStore Thrift服务端口 | 填写连接MetaStore的端口，默认为9083。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| 启用元数据仓库连接            | <p>在增量数据校验中，如果通过Metastore查询，当分区数量超过30,000时，可能会因为将所有分区信息加载到内存中而导致内存溢出（OOM）。因此，使用MySQL作为元数据仓库可以有效地解决这一问题。</p> <ul style="list-style-type: none"> <li>• 不启用：系统通过MetaStore查询Hive表的分区信息。</li> <li>• 启用：配置MySQL作为元数据仓库，系统将通过MySQL查询Hive表的分区信息。启用后，需要配置如下参数：                             <ul style="list-style-type: none"> <li>- 元仓类型：目前仅支持MySQL作为元数据仓库。</li> <li>- MySQL凭证：选择访问MySQL数据库的认证信息。认证信息需要添加到MgC Agent然后同步到MgC，方法请参考<a href="#">添加资源凭证</a>。</li> <li>- MySQL节点IP：输入MySQL数据库服务器的IP地址。</li> <li>- MySQL服务端口：输入MySQL数据库服务的端口号。</li> <li>- 数据库名：输入存储Hive表元数据的数据库名称。</li> </ul> </li> </ul> <p><b>说明</b><br/>请确保填写的MySQL凭证、节点IP、服务端口和数据库名与Hive实际使用的MySQL数据库相匹配，否则会导致数据校验失败。</p> |

表 1-78 Delta Lake（有元数据）连接参数配置说明

| 参数  | 配置说明  |
|-----|-------|
| 所属源 | 选择源端。 |

| 参数         | 配置说明                                                                                       |
|------------|--------------------------------------------------------------------------------------------|
| 连接名称       | 默认创建名称为“Delta-Lake-有元数据-4位随机数（字母和数字）”，也可以自定义名称。                                            |
| 执行机凭证      | 选择执行机的登录凭证。凭证的添加方法请参考 <a href="#">添加资源凭证</a> 中的大数据-执行机。                                    |
| 执行机IP      | 填写连接执行机的IP地址。                                                                              |
| 执行机连接端口    | 填写连接执行机的端口。                                                                                |
| Spark客户端目录 | 填写Spark客户端的安装目录路径。                                                                         |
| 环境变量地址     | 填写环境变量文件（配置文件）的绝对路径，例如：/opt/bigdata/client/bigdata_env。                                    |
| SQL文件地址    | 填写用于存放执行作业时生成的SQL文件的文件夹地址。您需要拥有该文件夹的读写权限。<br><b>须知</b><br>迁移完成后，您需要手动清理该路径下生成的文件夹，以释放存储空间。 |

表 1-79 Delta Lake（无元数据）连接参数配置说明

| 参数         | 配置说明                                                                                       |
|------------|--------------------------------------------------------------------------------------------|
| 所属源        | 选择源端。                                                                                      |
| 连接名称       | 默认创建名称为“Delta-Lake-无元数据-4位随机数（字母和数字）”，也可以自定义名称。                                            |
| 执行机凭证      | 选择执行机的登录凭证。凭证的添加方法请参考 <a href="#">添加资源凭证</a> 中的大数据-执行机。                                    |
| 执行机IP      | 填写连接执行机的IP地址。                                                                              |
| 执行机连接端口    | 填写连接执行机的端口。                                                                                |
| Spark客户端目录 | 填写Spark客户端的安装目录路径。                                                                         |
| 环境变量地址     | 填写环境变量文件（配置文件）的绝对路径，例如：/opt/bigdata/client/bigdata_env。                                    |
| SQL文件地址    | 填写用于存放执行作业时生成的SQL文件的文件夹地址。您需要拥有该文件夹的读写权限。<br><b>须知</b><br>迁移完成后，您需要手动清理该路径下生成的文件夹，以释放存储空间。 |

| 参数       | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 采集资源使用信息 | <p>可选参数。开启后，通过该连接创建的统计任务将收集大数据服务的资源使用信息，这些信息将用于MgC上的报告生成，以便于进行性能优化。</p> <p><b>须知</b><br/>使用该功能，需要MgC Agent添加的华为云账号AK/SK拥有MRS服务和DLI服务的只读权限（ReadOnly）。</p> <ul style="list-style-type: none"> <li>如果所选凭证与登录迁移中心 MgC的账号凭证相同，请勾选“<b>所选凭证与登录迁移中心凭证相同</b>”，系统将根据所选的区域，自动查询账号在该区域下的项目。 <ul style="list-style-type: none"> <li>区域：选择需要校验的数据所在区域/项目。</li> <li>项目：选择需要校验的数据所在项目。</li> <li>集群ID：输入需要校验的数据所在集群ID。</li> </ul> </li> <li>如果所选凭证与登录迁移中心 MgC的账号凭证不同，请手动输入如下参数信息： <ul style="list-style-type: none"> <li>区域ID：输入需要校验的数据所在区域ID。例如，区域为“华南-广州”，则输入“cn-south-1”</li> <li>项目ID：输入与区域对应的项目ID。</li> <li>集群ID：输入需要校验的数据所在集群ID。</li> </ul> </li> </ul> <p><b>说明</b></p> <ul style="list-style-type: none"> <li>区域ID和项目ID可以在“我的凭证 &gt; <a href="#">API凭证</a>”查看。</li> <li>集群ID的获取方法请参考获取<a href="#">MRS集群ID</a>。</li> </ul> |

表 1-80 Hudi（有元数据）连接参数配置说明

| 参数         | 配置说明                                                                                                   |
|------------|--------------------------------------------------------------------------------------------------------|
| 所属源        | 选择源端。                                                                                                  |
| 连接名称       | 默认创建名称为“Hudi-有元数据-4位随机数（字母和数字）”，也可以自定义名称。                                                              |
| 执行机凭证      | 选择执行机的登录凭证。凭证的添加方法请参考 <a href="#">添加资源凭证</a> 中的大数据-执行机。                                                |
| 执行机IP      | 填写连接执行机的IP地址。                                                                                          |
| 执行机连接端口    | 填写连接执行机的端口。                                                                                            |
| Spark客户端目录 | 填写Spark客户端的安装目录路径。                                                                                     |
| 环境变量地址     | 填写环境变量文件（配置文件）的绝对路径，例如：/opt/bigdata/client/bigdata_env。                                                |
| SQL文件地址    | <p>填写用于存放执行作业时生成的SQL文件的文件夹地址。您需要拥有该文件夹的读写权限。</p> <p><b>须知</b><br/>迁移完成后，您需要手动清理该路径下生成的文件夹，以释放存储空间。</p> |

表 1-81 Hudi（无元数据）连接参数配置说明

| 参数         | 配置说明                                                                                       |
|------------|--------------------------------------------------------------------------------------------|
| 所属源        | 选择源端。                                                                                      |
| 连接名称       | 默认创建名称为“Hudi-无元数据-4位随机数（字母和数字）”，也可以自定义名称。                                                  |
| 执行机凭证      | 选择执行机的登录凭证。凭证的添加方法请参考 <a href="#">添加资源凭证</a> 中的大数据-执行机。                                    |
| 执行机IP      | 填写连接执行机的IP地址。                                                                              |
| 执行机连接端口    | 填写连接执行机的端口。                                                                                |
| Spark客户端目录 | 填写Spark客户端的安装目录路径。                                                                         |
| 环境变量地址     | 填写环境变量文件（配置文件）的绝对路径，例如：/opt/bigdata/client/bigdata_env。                                    |
| SQL文件地址    | 填写用于存放执行作业时生成的SQL文件的文件夹地址。您需要拥有该文件夹的读写权限。<br><b>须知</b><br>迁移完成后，您需要手动清理该路径下生成的文件夹，以释放存储空间。 |

**步骤8** 参数配置完成后，单击“**测试连接**”，系统会校验配置信息并与组件建立连接，测试通过代表与组件连接成功。

**步骤9** 测试连接通过后，单击“**确认**”按钮，创建大数据组件连接完成。

**步骤10** 在**连接管理**页面，可以查看创建的连接列表以及基本信息。单击操作列的“**修改连接**”，可以修改连接参数配置。

----结束

## 1.12.4 创建表组并添加数据表

将源端数据表加入表组中，为后续对数作业做准备。

### 使用须知

- 每次最多支持导入10,000张表。
- 待导入的表必须属于同一个元数据来源。
- 导入数据表时，系统不会校验数据表加入的分组数量和对数标准。为保持对数的准确性，建议同一数据表加入的分组不超过3个，并避免加入对数标准相同的分组。
- 在创建用于MaxCompute迁移到DLI数据校验的表组时，建议先创建**count**、**sum**、**allsum**等基本校验标准的表组，在确保基本校验任务的校验标准已达标，并且源端和目的端数据量一致后，再创建**content**校验标准的表组进行详细的内容校验，以确保的目的端数据与源数据在数据量和内容上一致。

## 前提条件

- 已完成[准备工作](#)。
- 已[创建源端连接](#)。

## 创建表组

- 步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择创建的[大数据迁移项目](#)。
- 步骤2** 在左侧导航栏选择“[实施>大数据校验](#)”，进入[大数据校验](#)页面。
- 步骤3** 在功能入口区域，单击“[表管理](#)”，进入表管理页面的[表组](#)页签。
- 步骤4** 单击“[新增表组](#)”，弹出[新增表组](#)窗口。
- 步骤5** 根据[表1-82](#)，配置参数。

表 1-82 新增表组参数说明

| 参数     | 说明                                                                                               |
|--------|--------------------------------------------------------------------------------------------------|
| 表组名    | 用户自定义                                                                                            |
| 元数据连接  | 根据实际需求，选择 <a href="#">创建的源端连接</a> 。<br><b>注意</b><br>表组中的所有表必须来源于同一个元数据来源。                        |
| 校验标准   | 选择可以接受的数据差异最大值（即错误容忍度）。MgC提供了多种类型的对数标准可供选择，关于对数标准的详细说明，可以单击新增表组弹窗中的“ <a href="#">查看更多</a> ”进行查看。 |
| 说明（可选） | 您可以输入用于区分表组的说明。                                                                                  |

- 步骤6** 参数配置完成后，单击“[确认](#)”按钮，创建表组完成。在表组列表，可以查看创建的表组信息。

表组创建完成后，即可[导入数据表并加入表组](#)。

----结束

## 导入数据表并加入表组

- 步骤1** 在[表管理](#)页面，选择[库表清单](#)页签。
- 步骤2** 单击列表上方的“[表管理 > 导入数据表](#)”按钮，弹出[导入数据表](#)窗口。
- 步骤3** 选择元数据连接和要加入的表组名。
- 步骤4** 单击“[模板下载](#)”，将导入模板下载到本地。打开导入模板，将需要导入的数据表对应参数填写到模板内。

### 须知

- 每个模板最多支持导入10,000张表。
- 同一表组内的表必须属于同一个元数据来源。
- 模板中单元格不允许存在公式且必须是文本格式，否则会导致解析失败。
- 元数据连接为Delta（有元数据）或Hudi（有元数据）连接时，模板中的source\_path参数为必填项。
- 元数据连接为Delta（无元数据）或Hudi（无元数据）连接时，模板中的source\_path和target\_path参数为必填项。

**步骤5** 模板参数填写完成并保存后，单击导入窗口的“添加文件”按钮，将填写后的模板文件上传。

**步骤6** 上传成功后，单击“确认”按钮，系统开始导入。在库表清单列表中可以查看导入的数据表。

----结束

## 导出数据表

支持将Delta Lake和Hudi有元数据的数据表导出为CSV格式。无元数据的数据表不支持导出。

**步骤1** 在表管理页面，选择库表清单页签。

**步骤2** 单击列表上方的“导出 > 导出数据表”按钮，弹出导出窗口。



**步骤3** 选择需要导出的数据表所在表组，单击“确认”。

**步骤4** 导出完成后，单击“导出 > 导出管理”按钮，右侧弹出导出管理窗口。

**步骤5** 单击操作列的“下载”，可以将导出的数据表下载到本地。

----结束

### 1.12.5 创建目的端连接

### 1.12.5.1 创建执行机连接

执行机是获取数据的任务执行单元，需要分别与源端和目的端的执行机建立连接，然后执行数据统计，统计结果将用于数据校验。

以下大数据组件需要通过创建执行机连接来进行数据校验：

- Hive Metastore
- Delta Lake（有元数据）
- Delta Lake（无元数据）
- Hudi（有元数据）
- Hudi（无元数据）

#### ⚠ 注意

- 需要确保源端和目的端执行机都安装了大数据组件的命令行客户端。
- 源端和目的端执行机均需要创建连接。

## 操作步骤

- 步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择创建的[大数据迁移项目](#)。
- 步骤2** 在左侧导航栏选择“[实施>大数据校验](#)”，进入[大数据校验](#)页面。
- 步骤3** 在功能入口区域，单击“[迁移准备](#)”，进入[迁移准备](#)页面。
- 步骤4** 选择“[连接管理](#)”，单击“[创建连接](#)”，右侧弹出[新增连接](#)窗口。

图 1-61 创建连接



- 步骤5** 根据实际情况选择执行机类型，单击“[下一步](#)”，根据[表1-83](#)，配置执行机参数。

表 1-83 执行机参数配置说明

| 参数        | 配置说明                                                  |
|-----------|-------------------------------------------------------|
| 所属源       | 根据实际情况选择。<br><b>须知</b><br>CDH执行机只支持源端，MRS执行机支持源端和目的端。 |
| 连接名称      | 默认创建名称为“执行机类型-4位随机数（字母和数字）”，也可以自定义名称。                 |
| 迁移中心Agent | 选择在源端环境中安装的 <a href="#">MgC Agent（原Edge）</a> 。        |

| 参数      | 配置说明                                                                                       |
|---------|--------------------------------------------------------------------------------------------|
| 执行机凭证   | 选择执行机的登录凭证。凭证的添加方法请参考 <a href="#">添加资源凭证</a> 中的大数据-执行机。                                    |
| 执行机IP   | 填写连接执行机的IP地址。                                                                              |
| 执行机连接端口 | 填写连接执行机的端口，默认为22端口。                                                                        |
| 安装目录    | 填写MRS/CDH客户端的安装目录路径。即安装客户端时，“./install.sh”所填写的地址。                                          |
| SQL文件地址 | 填写用于存放执行作业时生成的SQL文件的文件夹地址。您需要拥有该文件夹的读写权限。<br><b>须知</b><br>校验完成后，您需要手动清理该路径下生成的文件夹，以释放存储空间。 |

- 步骤6** 参数配置完成后，单击“**测试连接**”，系统会校验配置信息并与执行机建立连接，测试通过代表与执行机连接成功。
- 步骤7** 测试连接通过后，单击“**确认**”按钮，创建执行机连接完成。
- 步骤8** 在**连接管理**页面，可以查看创建的执行机连接列表以及基本信息。单击操作列的“**修改连接**”，可以修改连接参数配置。

----结束

### 1.12.5.2 创建目的端云服务连接

对大数据云服务数据进行一致性校验时，需要分别与源端和目的端云服务建立连接，用于数据查询和计算。

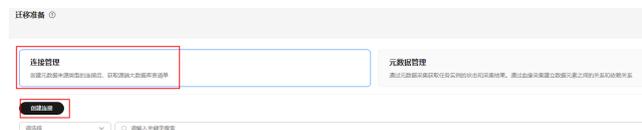
目前支持创建目的端连接的大数据云服务包括：

- 数据湖探索 ( DLI )
- CloudTable(ClickHouse)
- CloudTable(HBase)

### 操作步骤

- 步骤1** 登录**迁移中心管理控制台**。在左侧导航栏的项目下拉列表中选择创建的**大数据迁移项目**。
- 步骤2** 在左侧导航栏选择“**实施>大数据校验**”，进入**大数据校验**页面。
- 步骤3** 在**功能入口**区域，单击“**迁移准备**”，进入**迁移准备**页面。
- 步骤4** 选择“**连接管理**”，单击“**创建连接**”，右侧弹出**新增连接**窗口。

图 1-62 创建连接



**步骤5** 根据实际情况选择大数据云服务，单击“下一步”，进入详细配置页面。

**步骤6** 根据对应的参数配置说明，配置云服务参数。

- [DLI服务参数配置说明](#)
- [CloudTable\(ClickHouse\)服务参数配置说明](#)
- [CloudTable\(HBase\)服务参数配置说明](#)

**表 1-84** DLI 服务参数配置说明

| 参数        | 配置说明                                                                                                                                                     |
|-----------|----------------------------------------------------------------------------------------------------------------------------------------------------------|
| 所属源       | 选择目的端。                                                                                                                                                   |
| 连接名称      | 默认创建名称为“DLI-4位随机数（字母和数字）”，也可以自定义名称。                                                                                                                      |
| 迁移中心Agent | 选择在源端环境中安装的 <b>MgC Agent（原Edge）</b> 。                                                                                                                    |
| DLI凭证     | 选择DLI云服务的凭证。凭证的添加方法请参考 <a href="#">添加资源凭证</a> 中的大数据-DLI。如果所选凭证与登录迁移中心账号凭证相同，请勾选“ <b>所选凭证与登录迁移中心凭证相同</b> ”，系统将根据所选的区域，自动查询账号在该区域下的项目。                     |
| 区域ID      | 输入需要校验的数据（即目的端数据）所在区域的代码。例如：ap-southeast-1。可以参考 <a href="#">终端节点</a> 获取不同区域的区域代码信息。                                                                      |
| 项目ID      | 输入需要校验的数据所在的项目ID。获取方法请参考 <a href="#">获取项目的名称和ID</a> 。                                                                                                    |
| 队列名称      | 输入需要进行校验的DLI队列名称。队列类型必须是SQL队列。                                                                                                                           |
| 采集资源使用信息  | <p>可选参数。开启后，通过该连接创建的统计任务将收集大数据服务的资源使用信息，这些信息将用于MgC上的报告生成，以便于进行性能优化。</p> <p><b>须知</b><br/>使用该功能，需要MgC Agent添加的华为云账号AK/SK拥有MRS服务和DLI服务的只读权限（ReadOnly）。</p> |

**表 1-85** CloudTable(ClickHouse)服务参数配置说明

| 参数        | 配置说明                                                  |
|-----------|-------------------------------------------------------|
| 所属源       | 选择目的端。                                                |
| 连接名称      | 默认创建名称为“CloudTable-ClickHouse-4位随机数（字母和数字）”，也可以自定义名称。 |
| 迁移中心Agent | 选择在源端环境中安装的 <b>MgC Agent（原Edge）</b> 。                 |

| 参数                       | 配置说明                                                                                                                                                 |
|--------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------|
| CloudTable(ClickHouse)凭证 | 选择CloudTable(ClickHouse)集群的凭证。凭证的添加方法请参考 <a href="#">添加资源凭证</a> 中的大数据-ClickHouse。认证方式选择“ <b>用户名/密码</b> ”，填写在CloudTable服务创建的目的端ClickHouse集群的登录用户名和密码。 |
| 数据库链接                    | 填写连接到CloudTable(ClickHouse)的数据库链接地址。可以在ClickHouse集群的 <a href="#">基本信息</a> 获取，对应参数为“ <b>集群访问地址</b> ”。                                                 |

表 1-86 CloudTable(HBase)服务参数配置说明

| 参数                  | 配置说明                                                                                           |
|---------------------|------------------------------------------------------------------------------------------------|
| 所属源                 | 选择目的端。                                                                                         |
| 连接名称                | 默认创建名称为“CloudTable-HBase-4位随机数（字母和数字）”，也可以自定义名称。                                               |
| 迁移中心Agent           | 选择在源端环境中安装的 <a href="#">MgC Agent（原Edge）</a> 。                                                 |
| CloudTable(HBase)凭证 | 非安全集群，无需凭证。                                                                                    |
| 安全集群                | 当前仅支持非安全集群（普通集群）。                                                                              |
| 数据库链接               | 填写连接到CloudTable(HBase)的数据库链接地址。可以在HBase集群的 <a href="#">基本信息</a> 获取，对应参数为“ <b>ZK链接地址（内网）</b> ”。 |

**步骤7** 参数配置完成后，单击“**测试连接**”，系统会校验配置信息并与云服务建立连接，测试通过代表与云服务连接成功。

**步骤8** 测试连接通过后，单击“**确认**”按钮，创建云服务连接完成。

**步骤9** 在[连接管理](#)页面，可以查看创建的连接列表以及基本信息。单击操作列的“**修改连接**”，可以修改连接参数配置。

----结束

### 1.12.5.3 创建目的端组件连接

对大数据组件数据进行一致性校验时，需要与大数据组件建立连接，用于数据查询。

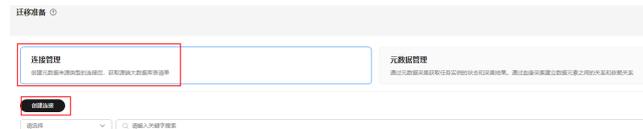
目前支持创建连接的大数据组件包括：

- Doris
- HBase
- ClickHouse
- Hive Metastore

## 操作步骤

- 步骤1 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择创建的[大数据迁移项目](#)。
- 步骤2 在左侧导航栏选择“[实施>大数据校验](#)”，进入[大数据校验](#)页面。
- 步骤3 在功能入口区域，单击“[迁移准备](#)”，进入[迁移准备](#)页面。
- 步骤4 选择“[连接管理](#)”，单击“[创建连接](#)”，右侧弹出[新增连接](#)窗口。

图 1-63 创建连接



- 步骤5 根据实际情况选择大数据组件，单击“[下一步](#)”，进入[详细配置](#)页面。
- 步骤6 根据对应的参数配置说明，配置组件连接参数。
  - [Doris组件参数配置说明](#)
  - [HBase组件参数配置说明](#)
  - [ClickHouse组件参数配置说明](#)
  - [Hive Metastore连接参数配置说明](#)

表 1-87 Doris 组件参数配置说明

| 参数        | 配置说明                                                                                                                                                         |
|-----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 所属源       | 选择目的端。                                                                                                                                                       |
| 连接名称      | 默认创建名称为“Doris-4位随机数（字母和数字）”，也可以自定义名称。                                                                                                                        |
| 迁移中心Agent | 选择在源端环境中安装的 <a href="#">MgC Agent（原Edge）</a> 。                                                                                                               |
| Doris凭证   | 选择在MgC Agent添加的目的端Doris凭证。凭证的添加方式请参考 <a href="#">添加资源凭证</a> ，类型选择“大数据-Doris”。                                                                                |
| 数据库地址     | 填写目的端Doris集群部署的服务器IP地址。<br>可登录 <a href="#">FusionInsight Manager</a> ，选择“ <a href="#">集群 &gt; 服务 &gt; Doris</a> ”，查看“ <a href="#">Leader所在的主机</a> ”获取。       |
| 数据库端口     | 填写目的端Doris集群部署的服务器的端口。<br>可登录 <a href="#">FusionInsight Manager</a> ，选择“ <a href="#">集群 &gt; 服务 &gt; Doris &gt; 配置</a> ”，搜索“ <a href="#">query_port</a> ”获取。 |
| 数据库名      | 填写目的端Doris数据库名称。                                                                                                                                             |

| 参数       | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 采集资源使用信息 | <p>可选参数。开启后，通过该连接创建的统计任务将收集大数据服务的资源使用信息，这些信息将用于MgC上的报告生成，以便于进行性能优化。</p> <p><b>须知</b><br/>使用该功能，需要MgC Agent添加的华为云账号AK/SK拥有MRS服务和DLI服务的只读权限（ReadOnly）。</p> <ul style="list-style-type: none"> <li>如果所选凭证与登录迁移中心 MgC的账号凭证相同，请勾选“<b>所选凭证与登录迁移中心凭证相同</b>”，系统将根据所选的区域，自动查询账号在该区域下的项目。 <ul style="list-style-type: none"> <li>区域：选择需要校验的数据所在区域/项目。</li> <li>项目：选择需要校验的数据所在项目。</li> <li>集群ID：输入需要校验的数据所在集群ID。</li> </ul> </li> <li>如果所选凭证与登录迁移中心 MgC的账号凭证不同，请手动输入如下参数信息： <ul style="list-style-type: none"> <li>区域ID：输入需要校验的数据所在区域ID。例如，区域为“华南-广州”，则输入“cn-south-1”</li> <li>项目ID：输入与区域对应的项目ID。</li> <li>集群ID：输入需要校验的数据所在集群ID。</li> </ul> </li> </ul> <p><b>说明</b></p> <ul style="list-style-type: none"> <li>区域ID和项目ID可以在“我的凭证 &gt; <b>API凭证</b>”查看。</li> <li>集群ID的获取方法请参考获取<b>MRS集群ID</b>。</li> </ul> |

表 1-88 HBase 组件参数配置说明

| 参数            | 配置说明                                                                 |
|---------------|----------------------------------------------------------------------|
| 所属源           | 选择目的端。                                                               |
| 连接名称          | 默认创建名称为“HBase-4位随机数（字母和数字）”，也可以自定义名称。                                |
| 迁移中心Agent     | 选择在源端环境中安装的 <b>MgC Agent（原Edge）</b> 。                                |
| HBase凭证       | 选择在MgC Agent添加的目的端HBase凭证。凭证的添加方式请参考 <b>添加资源凭证</b> ，类型选择“大数据-HBase”。 |
| 安全集群          | 选择是否部署在安全集群下。                                                        |
| Zookeeper节点IP | 填写连接到目的端Zookeeper的IP地址。通常是部署Zookeeper的服务器公网IP地址或内网IP地址。              |
| Zookeeper连接端口 | 填写连接到目的端Zookeeper的端口。默认为2181。                                        |
| HBase 版本      | 选择目的端HBase版本。                                                        |

| 参数       | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 采集资源使用信息 | <p>可选参数。开启后，通过该连接创建的统计任务将收集大数据服务的资源使用信息，这些信息将用于MgC上的报告生成，以便于进行性能优化。</p> <p><b>须知</b></p> <p>使用该功能，需要MgC Agent添加的华为云账号AK/SK拥有MRS服务和DLI服务的只读权限（ReadOnly）。</p> <ul style="list-style-type: none"> <li>如果所选凭证与登录迁移中心 MgC的账号凭证相同，请勾选“<b>所选凭证与登录迁移中心凭证相同</b>”，系统将根据所选的区域，自动查询账号在该区域下的项目。 <ul style="list-style-type: none"> <li>区域：选择需要校验的数据所在区域/项目。</li> <li>项目：选择需要校验的数据所在项目。</li> <li>集群ID：输入需要校验的数据所在集群ID。</li> </ul> </li> <li>如果所选凭证与登录迁移中心 MgC的账号凭证不同，请手动输入如下参数信息： <ul style="list-style-type: none"> <li>区域ID：输入需要校验的数据所在区域ID。例如，区域为“华南-广州”，则输入“cn-south-1”</li> <li>项目ID：输入与区域对应的项目ID。</li> <li>集群ID：输入需要校验的数据所在集群ID。</li> </ul> </li> </ul> <p><b>说明</b></p> <ul style="list-style-type: none"> <li>区域ID和项目ID可以在“我的凭证 &gt; <b>API凭证</b>”查看。</li> <li>集群ID的获取方法请参考获取<b>MRS集群ID</b>。</li> </ul> |

表 1-89 ClickHouse 组件参数配置说明

| 参数                   | 配置说明                                                                      |
|----------------------|---------------------------------------------------------------------------|
| 所属源                  | 选择目的端。                                                                    |
| 连接名称                 | 默认创建名称为“ClickHouse-4位随机数（字母和数字）”，也可以自定义名称。                                |
| 迁移中心Agent            | 选择在源端环境中安装的 <b>MgC Agent（原Edge）</b> 。                                     |
| ClickHouse凭证（可选）     | 选择在MgC Agent添加的目的端ClickHouse凭证。凭证的添加方法请参考 <b>添加资源凭证</b> 中的大数据-ClickHouse。 |
| 安全集群                 | 选择是否部署在安全集群下。                                                             |
| ClickHouse Server IP | 填写连接到目的端ClickHouse服务器的IP地址。通常是运行ClickHouse的主机IP地址。                        |
| HTTP端口               | 非安全集群下，填写与目的端ClickHouse 服务器通信的HTTP 端口。默认为8123。                            |
| HTTP SSL/TLS端口       | 安全集群下，填写与目的端ClickHouse 服务器通信的HTTPS 端口。                                    |

| 参数       | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 采集资源使用信息 | <p>可选参数。开启后，通过该连接创建的统计任务将收集大数据服务的资源使用信息，这些信息将用于MgC上的报告生成，以便于进行性能优化。</p> <p><b>须知</b><br/>使用该功能，需要MgC Agent添加的华为云账号AK/SK拥有MRS服务和DLI服务的只读权限（ReadOnly）。</p> <ul style="list-style-type: none"> <li>如果所选凭证与登录迁移中心 MgC的账号凭证相同，请勾选“<b>所选凭证与登录迁移中心凭证相同</b>”，系统将根据所选的区域，自动查询账号在该区域下的项目。 <ul style="list-style-type: none"> <li>区域：选择需要校验的数据所在区域/项目。</li> <li>项目：选择需要校验的数据所在项目。</li> <li>集群ID：输入需要校验的数据所在集群ID。</li> </ul> </li> <li>如果所选凭证与登录迁移中心 MgC的账号凭证不同，请手动输入如下参数信息： <ul style="list-style-type: none"> <li>区域ID：输入需要校验的数据所在区域ID。例如，区域为“华南-广州”，则输入“cn-south-1”</li> <li>项目ID：输入与区域对应的项目ID。</li> <li>集群ID：输入需要校验的数据所在集群ID。</li> </ul> </li> </ul> <p><b>说明</b></p> <ul style="list-style-type: none"> <li>区域ID和项目ID可以在“我的凭证 &gt; <b>API凭证</b>”查看。</li> <li>集群ID的获取方法请参考获取<b>MRS集群ID</b>。</li> </ul> |

表 1-90 Hive Metastore 连接参数配置说明

| 参数                   | 配置说明                                                                                                                                                                                     |
|----------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 所属源                  | 选择目的端。                                                                                                                                                                                   |
| 连接名称                 | 默认创建名称为“Hive-Metastore-4位随机数（字母和数字）”，也可以自定义名称。                                                                                                                                           |
| 迁移中心Agent            | 选择在源端环境中安装的 <b>MgC Agent（原Edge）</b> 。                                                                                                                                                    |
| 启用安全连接               | <p>选择是否启用安全连接。</p> <ul style="list-style-type: none"> <li>MetaStore在非安全集群下，可以不启用安全连接。</li> <li>MetaStore在安全集群下，需要启用安全连接并选择凭证。凭证的获取以及添加方法请参考<b>添加资源凭证</b>中的大数据-Hive Metastore。</li> </ul> |
| Hive版本               | <p>选择目的端Hive的实际版本。</p> <p><b>注意</b><br/>目前只支持3.x版本。</p>                                                                                                                                  |
| MetaStore节点IP        | 填写连接MetaStore的IP地址。                                                                                                                                                                      |
| MetaStore Thrift服务端口 | 填写连接MetaStore的端口，默认为9083。                                                                                                                                                                |

| 参数        | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|-----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 启用元数据仓库连接 | <p>在增量数据校验中，如果通过Metastore查询，当分区数量超过30,000时，可能会因为将所有分区信息加载到内存中而导致内存溢出（OOM）。因此，使用MySQL作为元数据仓库可以有效地解决这一问题。</p> <ul style="list-style-type: none"> <li>不启用：系统通过MetaStore查询Hive表的分区信息。</li> <li>启用：配置MySQL作为元数据仓库，系统将通过MySQL查询Hive表的分区信息。启用后，需要配置如下参数： <ul style="list-style-type: none"> <li>元仓类型：目前仅支持MySQL作为元数据仓库。</li> <li>MySQL凭证：选择访问MySQL数据库的认证信息。认证信息需要添加到MgC Agent然后同步到MgC，方法请参考<a href="#">添加资源凭证</a>。</li> <li>MySQL节点IP：输入MySQL数据库服务器的IP地址。</li> <li>MySQL服务端口：输入MySQL数据库服务的端口号。</li> <li>数据库名：输入存储Hive表元数据的数据库名称。</li> </ul> </li> </ul> <p><b>说明</b><br/>请确保填写的MySQL凭证、节点IP、服务端口和数据库名与Hive实际使用的MySQL数据库相匹配，否则会导致数据校验失败。</p> |

**步骤7** 参数配置完成后，单击“**测试连接**”，系统会校验配置信息并与组件建立连接，测试通过代表与组件连接成功。

**步骤8** 测试连接通过后，单击“**确认**”按钮，创建大数据组件连接完成。

**步骤9** 在**连接管理**页面，可以查看创建的连接列表以及基本信息。单击操作列的“**修改连接**”，可以修改连接参数配置。

----结束

## 1.12.6 创建统计任务并执行校验

利用创建的源端连接和目的端连接，分别创建源端和目的端统计任务，获取源端和目的端的数据信息，以便于进行数据校验。

支持的大数据组件和统计方式说明请参见[大数据校验概述](#)。

### 注意事项

- 创建的源端统计任务和目的端统计任务的统计方式必须保持一致。
- HBase安全集群任务与非安全集群任务不能同时执行。HBase安全集群需要加载认证信息，非安全集群连接时需要清除安全集群的认证信息，因此不能同时执行。
- 当源端服务（包含但不限于Lindorm、HBase）因欠费被锁定时，虽然可以正常创建数据连接和校验任务，但由于服务锁定限制了数据访问和操作权限，最终会导致校验任务无法正常执行。在进行数据校验前，请确保源端服务未因欠费被锁定，以避免权限问题。如果服务被锁定，请及时补缴费用以解除锁定。确认服务恢复正常后，重新尝试执行数据校验任务。

- 在进行Hive 2.x与Hive 3.x版本间数据校验时，由于Hive 2.x版本在查询固定长度字符串类型char(N)数据时，如果实际数据长度不满足N长度，Hive会以空格补齐至N长度。而Hive 3.x版本在查询时不会进行这种补齐操作。这可能导致在不同版本间比较结果时出现差异。为了避免因版本差异导致的数据不一致问题，建议使用Beeline作为客户端工具进行作业对数。
- 在完成MRS到MRS集群数据迁移后，如果源端和目的端都使用YARN模式进行校验，需要分开执行校验任务，确保一个任务执行完成后才能执行另一个任务。
- 在进行对数任务时，不支持使用 MRS 3.3.0 及以上版本集群节点的ECS作为执行机，会导致对数任务执行失败。

## 约束与限制

- 创建EMR Delta Lake到MRS Delta Lake数据校验任务存在以下约束与限制：
  - 源端EMR环境的Spark版本为3.3.1时，支持Delta Lake（有元数据）和Delta Lake（无元数据）的数据校验。
  - 源端EMR环境的Spark版本为2.4.8时，仅支持Delta Lake（有元数据）的数据校验。
- Hbase校验不支持纯冷数据表。
- 校验任务不支持跨天执行，执行过程中如果跨0点将导致校验结果异常，请合理规划任务以避免跨天运行。
- 在进行ClickHouse数据校验时，如果源端阿里云的ClickHouse版本为21.8.15.7，目的端华为云的ClickHouse版本为23.3.2.37，由于这两个版本在处理IPv4和IPv6字段类型以及相关函数计算结果上存在差异，因此无法进行字段校验。
- 在执行Hive数据的日级别增量、小时级增量以及指定日期的校验时，如果分区字段为date类型，且分区值的格式不是标准的YYYY-MM-DD时，系统目前不支持对这种场景进行数据校验。
- 不支持两端均为HBase 2.x不同版本的安全集群数据校验，因为不同版本间存在版本兼容性限制、安全认证机制差异、协议和接口不一致以及特性支持和配置不同等问题，都会影响校验结果。

## 前提条件

- 已[创建表组并添加数据表](#)。
- 已根据组件类型，[创建对应的连接](#)。

## 操作步骤

- 步骤1** 登录[迁移中心管理控制台](#)。在左侧导航栏的项目下拉列表中选择创建的[大数据迁移项目](#)。
- 步骤2** 在左侧导航栏选择“[实施>大数据校验](#)”，进入[大数据校验](#)页面。
- 步骤3** 在功能入口区域，单击“[任务管理](#)”，进入[任务管理](#)页面。
- 步骤4** 单击页面右上角的“[新增任务](#)”，右侧弹出[新增任务](#)窗口。
- 步骤5** 根据实际需求，选择大数据组件类型和统计方式，单击“[下一步](#)”，进入[详细配置](#)页面。
- 步骤6** 根据对应的参数配置说明，配置所选的统计方式参数。

 说明

各组件之间配置参数略有区别，请以实际页面为准。

- [全量对数作业参数配置说明](#)
- [日级增量校验参数配置说明](#)
- [小时级增量校验参数配置说明](#)
- [指定日期校验参数配置说明](#)
- [指定时间校验参数配置说明](#)

表 1-91 全量校验参数配置说明

| 区域   | 参数           | 配置说明                                                                                                                                                                                                                                   |
|------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 基础信息 | 任务名          | 默认创建名称为“组件类型全量校验-4位随机数（字母和数字）”，也可以自定义名称。                                                                                                                                                                                               |
| 任务配置 | 表组           | 选择需要校验的表所在表组。                                                                                                                                                                                                                          |
|      | 执行机连接        | Hive、Delta Lake和Hudi组件配置参数。 <ul style="list-style-type: none"> <li>• 创建源端统计任务时，选择源端执行机连接。</li> <li>• 创建目的端统计任务时，选择目的端执行机连接。</li> </ul>                                                                                                 |
|      | DLI连接        | 数据湖探索（DLI）组件配置参数，创建目的端统计任务。选择 <a href="#">创建的目的端DLI连接</a> 。                                                                                                                                                                            |
|      | Doris连接      | Doris组件配置参数。 <ul style="list-style-type: none"> <li>• 创建源端统计任务时，选择源端Doris连接。</li> <li>• 创建目的端统计任务时，选择目的端Doris连接。</li> </ul>                                                                                                            |
|      | HBase连接      | HBase组件和CloudTable(HBase)组件配置参数。 <ul style="list-style-type: none"> <li>• 创建源端统计任务时，选择源端HBase连接。</li> <li>• 创建目的端统计任务时，选择目的端HBase或CloudTable(HBase)连接。</li> </ul>                                                                      |
|      | ClickHouse连接 | ClickHouse、阿里云云数据库ClickHouse和CloudTable(Clickhouse)组件配置参数。 <ul style="list-style-type: none"> <li>• 创建源端统计任务时，选择源端MRS ClickHouse连接或阿里云云数据库ClickHouse连接。</li> <li>• 创建目的端统计任务时，选择目的端MRS ClickHouse或CloudTable(ClickHouse)连接。</li> </ul> |
|      | 元数据连接（可选）    | Delta Lake组件配置参数。<br>创建Delta Lake统计任务时，可以选择与表组所属集群相关的源端元数据连接，用于提升任务处理效率。                                                                                                                                                               |

| 区域 | 参数   | 配置说明                                                              |
|----|------|-------------------------------------------------------------------|
|    | 执行时间 | 设置作业自动执行时间。当作业激活后，每天都会在设置的时间自动执行。<br><b>说明</b><br>建议执行时间选择业务的低峰期。 |

| 区域 | 参数   | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|----|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    | 高级选项 | <ul style="list-style-type: none"> <li> <b>并发连接数</b>：进行数据统计时，单个执行机的任务并发数。默认为3，取值范围：1~10。<br/> <b>注意</b><br/>           在创建阿里云Hive校验任务时，请根据源端数据量和EMR Master节点的规格，合理设置并发连接数。建议按照以下两个条件进行设置：           <ul style="list-style-type: none"> <li>条件一：同时运行的任务并发数之和不超过元数据节点总核数的70%。</li> <li>条件二：任务分配的总资源不超过执行队列资源。计算如下：<br/>               内存：num-executors x executor-memory x 并发数<br/>               核数：num-executors x executor-cores x 并发数<br/>               例如，源端总数据量为500 G，且包含10000张表，其中8张为50 G、10万分区的大表。EMR Master节点规格为8U32G。             </li> </ul> <ul style="list-style-type: none"> <li>根据条件一：最大并发数 = <math>0.7 \times 8 = 5.6</math>，向下取整得到最大并发数为5。</li> <li>根据条件二：需要在后续执行命令参数配置中，选择<b>Spark-sql</b>，并设置如下参数：               <div data-bbox="901 963 1173 1153" style="border: 1px solid #ccc; padding: 5px; margin: 10px 0;"> <p>执行命令</p> <p>baseline hive <b>spark-sql</b></p> <p>executor-memory 4G</p> <p>master yarn</p> <p>num-executors 20</p> <p>executor-cores 2</p> <p>driver-memory 10G <a href="#">添加参数</a></p> <p>脚本预览</p> <pre>spark-sql --executor-memory 4G --master yarn --num-executors 20 --executor-cores 2 --driver-memory 10G -f</pre> </div> <ul style="list-style-type: none"> <li>executor-memory = 4G</li> <li>master = yarn</li> <li>num-executors = 20</li> <li>executor-cores = 2</li> <li>driver-memory = 10G</li> </ul> </li> </ul> </li> <li> <b>单文件SQL条数</b>：每个需要校验的表会产生一条SQL用以获取数据摘要，每批提交给集群处理的数据文件中所包含的SQL条数。默认为10，推荐取值范围：1~50。         </li> <li> <b>超时时间</b>：每个SQL语句在大数据集群上执行时的最大允许时长，单位为秒（s）。默认为600，推荐取值范围：600~ 7200。         </li> <li> <b>发送SMN通知</b>：选择是否使用SMN消息通知功能，该功能可以通过电子邮件、短信或者自定义URL等方式及时通知您任务状态。<br/> <b>须知</b> <ul style="list-style-type: none"> <li>使用该功能前，需要前往SMN控制台创建主题。方法请参考<a href="#">创建SMN主题</a>。</li> <li>使用该功能可能产生少量费用，具体费用由SMN服务结算。详情请查看<a href="#">SMN计费说明</a>。</li> </ul> </li> </ul> |

| 区域                               | 参数       | 配置说明                                                                                                                                                                                                                                                        |
|----------------------------------|----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 数据筛选                             | 非分区表处理策略 | <p>选择对于非分区表的校验处理策略。</p> <ul style="list-style-type: none"> <li>● <b>全部校验</b>：全量处理所有非分区表，不做筛选。</li> <li>● <b>全部忽略</b>：忽略所有非分区表，不做校验。</li> <li>● <b>按更新时间过滤</b>：只处理更新时间在校验范围内的非分区表。<b>注意</b>：非分区表非insert写入业务，元数据更新时间存在偏差，按更新时间过滤分区时可能遗漏分区，慎重选择该选项！</li> </ul> |
|                                  | 高级选项     | <p><b>单SQL字段上限</b>：每条SQL查询的表的字段上限。单SQL查询的字段数越多，执行时间越长，如果设置的字段数上限过少，则可能形成更多SQL，影响获取校验结果的效率。默认是0，即无上限，取值范围100~500。</p>                                                                                                                                        |
| OBS桶配置检查                         | -        | <ul style="list-style-type: none"> <li>● 如果需要将任务日志文件和内容校验结果上传到OBS桶进行管理和分析，请参考<a href="#">配置OBS桶</a>配置日志桶。配置了日志桶后，会默认使用日志上传功能，将日志数据和内容校验结果上传到指定的OBS桶中。</li> <li>● 如果不需要将任务日志文件和内容校验结果上传到OBS桶，请勾选“<b>我确认仅需在迁移中心Agent侧查看日志和内容校验结果，无需上传至OBS桶</b>”。</li> </ul>   |
| 命令参数<br>(Hive配置参数)               | 安全配置     | <p>如果大数据集群启用了安全配置（例如Kerberos安全认证），请勾选此选项，并配置安全认证命令。配置前，需要您手动将包含认证密钥的.keytab文件上传至执行机。</p> <ul style="list-style-type: none"> <li>● <b>keytab文件地址</b>：输入.keytab文件在执行机的存放地址。</li> <li>● <b>keytab文件用户名</b>：输入.keytab文件对应的用户名。</li> </ul>                       |
|                                  | 执行命令     | <p>可以通过设置Beeline或Spark-sql命令行参数，对数据进行处理。</p> <ul style="list-style-type: none"> <li>● <b>Beeline</b>：用于与Hive交互的命令行工具。</li> <li>● <b>Spark-sql</b>：用于执行SQL查询和分析数据。</li> </ul>                                                                                |
| MaxCompute参数<br>(MaxCompute配置参数) | -        | <p>根据实际需求，添加MaxCompute参数。参数详情请参考<a href="#">MaxCompute文档</a>。</p>                                                                                                                                                                                           |
| 统计配置<br>(DLI配置参数)                | 参数       | <p>根据实际需求进行配置，支持的自定义参数详情请参考<a href="#">自定义参数说明</a>。</p>                                                                                                                                                                                                     |

| 区域                              | 参数                               | 配置说明                                                                                                                                                                                                                                                     |
|---------------------------------|----------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 统计配置<br>( HBase配置参数 )           | 运行模式                             | 支持以下两种运行模式：<br>Yarn：适用于大型分布式环境，可以充分利用集群资源，提高任务的并行度和效率。<br>Local：适用于小规模数据集或开发测试环境，可以快速进行调试和验证。                                                                                                                                                            |
|                                 | 参数                               | 根据所选的运行模式和实际需求，添加命令行参数。                                                                                                                                                                                                                                  |
| 命令参数<br>( Delta Lake和Hudi配置参数 ) | 安全配置 ( 仅创建Delta Lake目的端任务时可以配置 ) | 如果大数据集群启用了安全配置 ( 例如Kerberos安全认证 )，请勾选此选项，并配置安全认证命令。配置前，需要您手动将包含认证密钥的 <b>.keytab</b> 文件上传至执行机。 <ul style="list-style-type: none"> <li>• <b>keytab文件地址</b>：输入<b>.keytab</b>文件在执行机的存放地址。</li> <li>• <b>keytab文件用户名</b>：输入<b>.keytab</b>文件对应的用户名。</li> </ul> |
|                                 | spark-sql命令                      | Spark SQL是Spark中用于结构化数据处理的模块。在Spark应用中，可以无缝地使用SQL语句亦或是DataSet API对结构化数据进行查询。更多介绍请参见 <a href="#">SparkSQL原理</a> 。建议保持默认。                                                                                                                                  |

| 区域 | 参数              | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|----|-----------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    | spark-submit 命令 | <p>通过Spark基本shell命令，提交Spark应用。命令为：</p> <pre>./bin/spark-submit \ --class &lt;main-class&gt; \ --master &lt;master-url&gt; \ ... # other options &lt;application-jar&gt; \ [application-arguments]</pre> <p>参数解释：</p> <ul style="list-style-type: none"> <li>● --class: Spark应用的类名。</li> <li>● --master: Spark用于所连接的master，如 yarn-client, yarn-cluster等。</li> <li>● application-jar: Spark应用的jar包的路径。</li> <li>● application-arguments: 提交Spark应用的所需要的参数（可以为空）。</li> </ul> <p>以下为特定场景下需要添加的参数配置：</p> <ul style="list-style-type: none"> <li>● 如果创建的是阿里云 EMR且涉及Spark 3的 Delta Lake源端统计任务，请添加如下参数： <ul style="list-style-type: none"> <li>- 参数名: jars</li> <li>- 参数值: '/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-core_2.12-*.jar,/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-storage-*.jar'</li> </ul> <p><b>注意</b><br/>参数值需要根据实际环境目录和Delta Lake 的版本进行替换。</p> </li> <li>● 如果源端是阿里云 EMR Delta 2.1.0版本并且配置了Spark 2.4.8来处理Delta Lake数据，请添加如下参数： <ul style="list-style-type: none"> <li>- 参数名: mgc.delta.spark.version</li> <li>- 参数值: 2</li> </ul> </li> </ul> |

表 1-92 日级增量校验参数配置说明

| 区域   | 参数    | 配置说明                                                                                                                                          |
|------|-------|-----------------------------------------------------------------------------------------------------------------------------------------------|
| 基础信息 | 任务名   | 默认创建名称为“组件类型日级增量校验-4位随机数（字母和数字）”，也可以自定义名称。                                                                                                    |
| 任务配置 | 表组    | 选择需要校验的表所在表组。                                                                                                                                 |
|      | 执行机连接 | <p>Hive、Delta Lake和Hudi组件配置参数。</p> <ul style="list-style-type: none"> <li>● 创建源端统计任务时，选择源端执行机连接。</li> <li>● 创建目的端统计任务时，选择目的端执行机连接。</li> </ul> |

| 区域 | 参数              | 配置说明                                                                                                                                        |
|----|-----------------|---------------------------------------------------------------------------------------------------------------------------------------------|
|    | DLI连接           | 数据湖探索 ( DLI ) 组件配置参数，创建目的端统计任务。选择 <a href="#">创建的DLI连接</a> 。                                                                                |
|    | Doris连接         | Doris组件配置参数。<br><ul style="list-style-type: none"> <li>创建源端统计任务时，选择源端Doris连接。</li> <li>创建目的端统计任务时，选择目的端Doris连接。</li> </ul>                  |
|    | HBase连接         | HBas组件配置参数。<br><ul style="list-style-type: none"> <li>创建源端统计任务时，选择源端HBase连接。</li> <li>创建目的端统计任务时，选择目的端HBase或CloudTable(HBase)连接。</li> </ul> |
|    | 元数据连接<br>( 可选 ) | Hive、Delta Lake组件配置参数。<br>创建Hive目的端统计任务时可配置，选择Hive目的端元数据连接。用于判断目的端分区是否存在。<br>创建Delta Lake统计任务时，可以选择与表组所属集群相关的源端元数据连接，用于提升任务处理效率。            |
|    | 执行时间            | 设置作业自动执行时间。当作业激活后，每天都会在设置的时间自动执行。                                                                                                           |

| 区域 | 参数   | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|----|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    | 高级选项 | <ul style="list-style-type: none"> <li> <b>并发连接数</b>：进行数据统计时，单个执行机的任务并发数。默认为3，取值范围：1~10。<br/> <b>注意</b><br/>                     在创建阿里云Hive校验任务时，请根据源端数据量和EMR Master节点的规格，合理设置并发连接数。建议按照以下两个条件进行设置：                     <ul style="list-style-type: none"> <li>条件一：同时运行的任务并发数之和不超过元数据节点总核数的70%。</li> <li>条件二：任务分配的总资源不超过执行队列资源。计算如下：<br/>                             内存：num-executors x executor-memory x 并发数<br/>                             核数：num-executors x executor-cores x 并发数<br/>                             例如，源端总数据量为500 G，且包含10000张表，其中8张为50 G、10万分区的大表。EMR Master节点规格为8U32G。                             <ul style="list-style-type: none"> <li>根据条件一：最大并发数 = <math>0.7 \times 8 = 5.6</math>，向下取整得到最大并发数为5。</li> <li>根据条件二：需要在后续执行命令参数配置中，选择<b>Spark-sql</b>，并设置如下参数：                                     <div data-bbox="895 958 1166 1144" style="border: 1px solid #ccc; padding: 5px; margin: 10px 0;"> <p>执行命令</p> <p>baseline hive <b>spark-sql</b></p> <p>executor-memory 4G</p> <p>master yarn</p> <p>num-executors 20</p> <p>executor-cores 2</p> <p>driver-memory 10G <span style="float: right;">添加参数</span></p> <p>脚本预览</p> <pre>spark-sql --executor-memory 4G --master yarn --num-executors 20 --executor-cores 2 --driver-memory 10G -f</pre> </div> <br/>                             executor-memory = 4G<br/>                             master = yarn<br/>                             num-executors = 20<br/>                             executor-cores = 2<br/>                             driver-memory = 10G                             <ul style="list-style-type: none"> <li> <b>单文件SQL条数</b>：每个需要校验的表会产生一条SQL用以获取数据摘要，每批提交给集群处理的数据文件中所包含的SQL条数。默认为10，推荐取值范围：1~50。</li> <li> <b>超时时间</b>：每个SQL语句在大数据集群上执行时的最大允许时长，单位为秒（s）。默认为600，推荐取值范围：600~ 7200。</li> <li> <b>发送SMN通知</b>：选择是否使用SMN消息通知功能，该功能可以通过电子邮件、短信或者自定义URL等方式及时通知您任务状态。<br/> <b>须知</b> <ul style="list-style-type: none"> <li>使用该功能前，需要前往SMN控制台创建主题。方法请参考<a href="#">创建SMN主题</a>。</li> <li>使用该功能可能产生少量费用，具体费用由SMN服务结算。详情请查看<a href="#">SMN计费说明</a>。</li> </ul> </li> </ul> </li> </ul> </li> </ul> </li></ul> |

| 区域       | 参数       | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|----------|----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 数据筛选     | 增量区间     | <p>选择需要对哪些时间段内的增量数据进行校验。默认统计24小时增量数据。T代表执行时间，T-n代表执行时间前的n x 24小时。</p> <p>勾选<b>统计连续日期</b>，可以选择多个统计时间。</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|          | 非分区表处理策略 | <p>选择对于非分区表的对数处理策略。</p> <ul style="list-style-type: none"> <li>● <b>全部校验</b>：全量处理所有非分区表，不做筛选。</li> <li>● <b>全部忽略</b>：忽略所有非分区表，不做校验。</li> <li>● <b>按更新时间过滤</b>：只处理更新时间在校验范围内的非分区表。<b>注意</b>：非分区表非insert写入业务，元数据更新时间存在偏差，按更新时间过滤分区时可能遗漏分区，慎重选择该选项！</li> </ul>                                                                                                                                                                                                                                                                                                         |
|          | 高级选项     | <ul style="list-style-type: none"> <li>● <b>分区筛选方式</b>：通过“更新时间”和“创建时间”，判断分区是否属于校验范围内。 <ul style="list-style-type: none"> <li>- <b>更新时间</b>：指分区内数据最后一次被修改或更新的时间戳。如果您关心的是分区内数据的最新状态或最近一段时间内的数据变更，可以选择“更新时间”作为筛选条件。</li> <li>- <b>创建时间</b>：指分区被首次创建时的时间戳。如果您关心的是分区从创建开始到某个时间点内的数据，可以选择“创建时间”作为筛选条件。</li> </ul> </li> <li>● <b>分区上限</b>：用于限制查询或处理的分区数量，避免不必要的全表扫描或处理大量不需要的数据。默认为3，取值范围：1~50。<br/>例如，设置分区上限为3，系统只处理降序排序后分区列表中的前3个分区。</li> <li>● <b>单SQL字段上限</b>：每条SQL查询的表的字段上限。单SQL查询的字段数越多，执行时间越长，如果设置的字段数上限过少，则可能形成更多SQL，影响获取校验结果的效率。默认是0，即无上限，取值范围100~500。</li> </ul> |
| OBS桶配置检查 | -        | <ul style="list-style-type: none"> <li>● 如果需要将任务日志文件和内容校验结果上传到OBS桶进行管理和分析，请参考<a href="#">配置OBS桶</a>配置日志桶。配置了日志桶后，会默认使用日志上传功能，将日志数据和内容校验结果上传到指定的OBS桶中。</li> <li>● 如果不需要将任务日志文件和内容校验结果上传到OBS桶，请勾选“<b>我确认仅需在迁移中心Agent侧查看日志和内容校验结果，无需上传至OBS桶</b>”。</li> </ul>                                                                                                                                                                                                                                                                                                           |

| 区域                               | 参数                            | 配置说明                                                                                                                                                                                                                              |
|----------------------------------|-------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 执行脚本<br>(Hive配置参数)               | 安全配置                          | <p>如果大数据集群启用了安全配置（例如Kerberos安全认证），请勾选此选项，并配置安全认证命令。配置前，需要您手动将包含认证密钥的.keytab文件上传至执行机。</p> <ul style="list-style-type: none"> <li><b>keytab文件地址</b>：输入.keytab文件在执行机的存放地址。</li> <li><b>keytab文件用户名</b>：输入.keytab文件对应的用户名。</li> </ul> |
|                                  | 执行命令                          | <p>可以通过设置Beeline或Spark-sql命令行参数，对数据进行处理。</p> <ul style="list-style-type: none"> <li><b>Beeline</b>：用于与Hive交互的命令行工具。</li> <li><b>Spark-sql</b>：用于执行SQL查询和分析数据。</li> </ul>                                                          |
| MaxCompute参数<br>(MaxCompute配置参数) | -                             | 根据实际需求，添加MaxCompute参数。参数详情请参考 <a href="#">MaxCompute文档</a> 。                                                                                                                                                                      |
| 统计配置<br>(DLI配置参数)                | 参数                            | 根据实际需求进行配置，支持的自定义参数详情请参考 <a href="#">自定义参数说明</a> 。                                                                                                                                                                                |
| 统计配置<br>(HBase配置参数)              | 运行模式                          | <p>支持以下两种运行模式：</p> <p>Yarn：适用于大型分布式环境，可以充分利用集群资源，提高任务的并行度和效率。</p> <p>Local：适用于小规模数据集或开发测试环境，可以快速进行调试和验证。</p>                                                                                                                      |
|                                  | 参数                            | 根据所选的运行模式和实际需求，添加命令行参数。                                                                                                                                                                                                           |
| 命令参数<br>(Delta Lake和Hudi配置参数)    | 安全配置（仅创建Delta Lake目的端任务时可以配置） | <p>如果大数据集群启用了安全配置（例如Kerberos安全认证），请勾选此选项，并配置安全认证命令。配置前，需要您手动将包含认证密钥的.keytab文件上传至执行机。</p> <ul style="list-style-type: none"> <li><b>keytab文件地址</b>：输入.keytab文件在执行机的存放地址。</li> <li><b>keytab文件用户名</b>：输入.keytab文件对应的用户名。</li> </ul> |
|                                  | spark-sql命令                   | Spark SQL是Spark中用于结构化数据处理的模块。在Spark应用中，可以无缝地使用SQL语句亦或是DataSet API对结构化数据进行查询。更多介绍请参见 <a href="#">SparkSQL原理</a> 。建议保持默认。                                                                                                           |

| 区域 | 参数              | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
|----|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    | spark-submit 命令 | <p>通过Spark基本shell命令，提交Spark应用。命令为：</p> <pre>./bin/spark-submit \ --class &lt;main-class&gt; \ --master &lt;master-url&gt; \ ... # other options &lt;application-jar&gt; \ [application-arguments]</pre> <p>参数解释：</p> <ul style="list-style-type: none"> <li>● --class: Spark应用的类名。</li> <li>● --master: Spark用于所连接的master，如yarn-client, yarn-cluster等。</li> <li>● application-jar: Spark应用的jar包的路径。</li> <li>● application-arguments: 提交Spark应用的所需要的参数（可以为空）。</li> </ul> <p>以下为特定场景下需要添加的参数配置：</p> <ul style="list-style-type: none"> <li>● 如果创建的是阿里云 EMR且涉及Spark 3的Delta Lake源端统计任务，请添加如下参数： <ul style="list-style-type: none"> <li>- 参数名: jars</li> <li>- 参数值: '/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-core_2.12-*.jar,/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-storage-*.jar'</li> </ul> <p><b>注意</b><br/>参数值需要根据实际环境目录和Delta Lake 的版本进行替换。</p> </li> <li>● 如果源端是阿里云 EMR Delta 2.1.0版本并且配置了Spark 2.4.8来处理Delta Lake数据，请添加如下参数： <ul style="list-style-type: none"> <li>- 参数名: mgc.delta.spark.version</li> <li>- 参数值: 2</li> </ul> </li> </ul> |

表 1-93 小时级增量校验参数配置说明

| 区域   | 参数    | 配置说明                                                                                                                                          |
|------|-------|-----------------------------------------------------------------------------------------------------------------------------------------------|
| 基础信息 | 任务名   | 默认创建名称为“组件类型小时级增量校验-4位随机数（字母和数字）”，也可以自定义名称。                                                                                                   |
| 任务配置 | 表组    | 选择需要校验的表所在表组。                                                                                                                                 |
|      | 执行机连接 | <p>Hive、Delta Lake和Hudi组件配置参数。</p> <ul style="list-style-type: none"> <li>● 创建源端统计任务时，选择源端执行机连接。</li> <li>● 创建目的端统计任务时，选择目的端执行机连接。</li> </ul> |

| 区域 | 参数        | 配置说明                                                                                                                             |
|----|-----------|----------------------------------------------------------------------------------------------------------------------------------|
|    | DLI连接     | 数据湖探索（DLI）组件配置参数，创建目的端统计任务。选择 <a href="#">创建的DLI连接</a> 。                                                                         |
|    | Doris连接   | Doris组件配置参数。<br><ul style="list-style-type: none"> <li>创建源端统计任务时，选择源端Doris连接。</li> <li>创建目的端统计任务时，选择目的端Doris连接。</li> </ul>       |
|    | 元数据连接（可选） | Hive、Delta Lake组件配置参数。<br>创建Hive目的端统计任务时可配置，选择Hive目的端元数据连接。用于判断目的端分区是否存在。<br>创建Delta Lake统计任务时，可以选择与表组所属集群相关的源端元数据连接，用于提升任务处理效率。 |
|    | 开始时间      | 设置作业自动执行时间。当作业激活后，每天都会在设置的时间自动执行。                                                                                                |

| 区域   | 参数   | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|------|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|      | 高级选项 | <ul style="list-style-type: none"> <li> <b>并发连接数</b>：进行数据统计时，单个执行机的任务并发数。默认为3，取值范围：1~10。<br/> <b>注意</b><br/>                     在创建阿里云Hive校验任务时，请根据源端数据量和EMR Master节点的规格，合理设置并发连接数。建议按照以下两个条件进行设置：                     <ul style="list-style-type: none"> <li>条件一：同时运行的任务并发数之和不超过元数据节点总核数的70%。</li> <li>条件二：任务分配的总资源不超过执行队列资源。计算如下：<br/>                             内存：num-executors x executor-memory x 并发数<br/>                             核数：num-executors x executor-cores x 并发数<br/>                             例如，源端总数据量为500 G，且包含10000张表，其中8张为50 G、10万分区的大表。EMR Master节点规格为8U32G。                             <ul style="list-style-type: none"> <li>根据条件一：最大并发数 = 0.7×8=5.6，向下取整得到最大并发数为5。</li> <li>根据条件二：需要在后续执行命令参数配置中，选择<b>Spark-sql</b>，并设置如下参数：</li> </ul> </li> </ul> </li> </ul> <div data-bbox="869 958 1273 1146" style="border: 1px solid #ccc; padding: 5px; margin: 10px 0;"> <p>执行命令</p> <p>baseline hive <b>spark-sql</b></p> <p>executor-memory 4G</p> <p>master yarn</p> <p>num-executors 20</p> <p>executor-cores 2</p> <p>driver-memory 10G <a href="#">添加参数</a></p> <p>脚本预览</p> <pre>spark-sql --executor-memory 4G --master yarn --num-executors 20 --executor-cores 2 --driver-memory 10G -f</pre> </div> <ul style="list-style-type: none"> <li> <b>单文件SQL条数</b>：每个需要校验的表会产生一条SQL用以获取数据摘要，每批提交给集群处理的数据文件中所包含的SQL条数。默认为10，推荐取值范围：1~50。                 </li> <li> <b>超时时间</b>：每个SQL语句在大数据集群上执行时的最大允许时长，单位为秒（s）。默认为600，推荐取值范围：600~ 7200。                 </li> <li> <b>发送SMN通知</b>：选择是否使用SMN消息通知功能，该功能可以通过电子邮件、短信或者自定义URL等方式及时通知您任务状态。<br/> <b>须知</b> <ul style="list-style-type: none"> <li>使用该功能前，需要前往SMN控制台创建主题。方法请参考<a href="#">创建SMN主题</a>。</li> <li>使用该功能可能产生少量费用，具体费用由SMN服务结算。详情请查看<a href="#">SMN计费说明</a>。</li> </ul> </li> </ul> |
| 数据筛选 | 执行周期 | 设置任务的执行频率和持续时间。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |

| 区域             | 参数       | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|----------------|----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                | 非分区表处理策略 | <p>选择对于非分区表的校验处理策略。</p> <ul style="list-style-type: none"> <li>● <b>全部校验</b>：全量处理所有非分区表，不做筛选。</li> <li>● <b>全部忽略</b>：忽略所有非分区表，不做校验。</li> <li>● <b>按更新时间过滤</b>：只处理更新时间在校验范围内的非分区表。<b>注意</b>：非分区表非insert写入业务，元数据更新时间存在偏差，按更新时间过滤分区时可能遗漏分区，慎重选择该选项！</li> </ul>                                                                                                                                                                                                                                                                                                         |
|                | 高级选项     | <ul style="list-style-type: none"> <li>● <b>分区筛选方式</b>：通过“更新时间”和“创建时间”，判断分区是否属于校验范围内。 <ul style="list-style-type: none"> <li>- <b>更新时间</b>：指分区内数据最后一次被修改或更新的时间戳。如果您关心的是分区内数据的最新状态或最近一段时间内的数据变更，可以选择“更新时间”作为筛选条件。</li> <li>- <b>创建时间</b>：指分区被首次创建时的时间戳。如果您关心的是分区从创建开始到某个时间点内的数据，可以选择“创建时间”作为筛选条件。</li> </ul> </li> <li>● <b>分区上限</b>：用于限制查询或处理的分区数量，避免不必要的全表扫描或处理大量不需要的数据。默认为3，取值范围：1~50。<br/>例如，设置分区上限为3，系统只处理降序排序后分区列表中的前3个分区。</li> <li>● <b>单SQL字段上限</b>：每条SQL查询的表的字段上限。单SQL查询的字段数越多，执行时间越长，如果设置的字段数上限过少，则可能形成更多SQL，影响获取校验结果的效率。默认是0，即无上限，取值范围100~500。</li> </ul> |
| OBS桶配置检查       | -        | <ul style="list-style-type: none"> <li>● 如果需要将任务日志文件和内容校验结果上传到OBS桶进行管理和分析，请参考<a href="#">配置OBS桶</a>配置日志桶。配置了日志桶后，会默认使用日志上传功能，将日志数据和内容校验结果上传到指定的OBS桶中。</li> <li>● 如果不需要将任务日志文件和内容校验结果上传到OBS桶，请勾选“<b>我确认仅需在迁移中心Agent侧查看日志和内容校验结果，无需上传至OBS桶</b>”。</li> </ul>                                                                                                                                                                                                                                                                                                           |
| 执行脚本（Hive配置参数） | 安全配置     | <p>如果大数据集群启用了安全配置（例如Kerberos安全认证），请勾选此选项，并配置安全认证命令。配置前，需要您手动将包含认证密钥的keytab文件上传至执行机。</p> <ul style="list-style-type: none"> <li>● <b>keytab文件地址</b>：输入keytab文件在执行机的存放地址。</li> <li>● <b>keytab文件用户名</b>：输入keytab文件对应的用户名。</li> </ul>                                                                                                                                                                                                                                                                                                                                  |

| 区域                                 | 参数                             | 配置说明                                                                                                                                                                                                                                  |
|------------------------------------|--------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                    | 执行命令                           | <p>可以通过设置Beeline或Spark-sql命令行参数，对数据进行处理。</p> <ul style="list-style-type: none"> <li>• <b>Beeline</b>: 用于与 Hive交互的命令行工具。</li> <li>• <b>Spark-sql</b>: 用于执行 SQL 查询和分析数据。</li> </ul>                                                     |
| MaxCompute参数<br>(MaxCompute组件配置参数) | -                              | 根据实际需求，添加MaxCompute参数。参数详情请参考 <a href="#">MaxCompute文档</a> 。                                                                                                                                                                          |
| 统计配置<br>(DLI配置参数)                  | 参数                             | 根据实际需求进行配置，支持的自定义参数详情请参考 <a href="#">自定义参数说明</a> 。                                                                                                                                                                                    |
| 命令参数<br>(Delta Lake和Hudi配置参数)      | 安全配置 (仅创建Delta Lake目的端任务时可以配置) | <p>如果大数据集群启用了安全配置 (例如Kerberos安全认证)，请勾选此选项，并配置安全认证命令。配置前，需要您手动将包含认证密钥的keytab文件上传至执行机。</p> <ul style="list-style-type: none"> <li>• <b>keytab文件地址</b>: 输入keytab文件在执行机的存放地址。</li> <li>• <b>keytab文件用户名</b>: 输入keytab文件对应的用户名。</li> </ul> |
|                                    | spark-sql命令                    | <p>Spark SQL是Spark中用于结构化数据处理的模块。在Spark应用中，可以无缝地使用SQL语句亦或是DataSet API对结构化数据进行查询。更多介绍请参见<a href="#">SparkSQL原理</a>。建议保持默认。</p>                                                                                                          |

| 区域 | 参数              | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
|----|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    | spark-submit 命令 | <p>通过Spark基本shell命令，提交Spark应用。命令为：</p> <pre>./bin/spark-submit \ --class &lt;main-class&gt; \ --master &lt;master-url&gt; \ ... # other options &lt;application-jar&gt; \ [application-arguments]</pre> <p>参数解释：</p> <ul style="list-style-type: none"> <li>• --class: Spark应用的类名。</li> <li>• --master: Spark用于所连接的master，如yarn-client, yarn-cluster等。</li> <li>• application-jar: Spark应用的jar包的路径。</li> <li>• application-arguments: 提交Spark应用的所需要的参数（可以为空）。</li> </ul> <p>以下为特定场景下需要添加的参数配置：</p> <ul style="list-style-type: none"> <li>• 如果创建的是阿里云 EMR且涉及Spark 3的Delta Lake源端统计任务，请添加如下参数： <ul style="list-style-type: none"> <li>- 参数名: jars</li> <li>- 参数值: '/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-core_2.12-*.jar,/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-storage-*.jar'</li> </ul> <p><b>注意</b><br/>参数值需要根据实际环境目录和Delta Lake 的版本进行替换。</p> </li> <li>• 如果源端是阿里云 EMR Delta 2.1.0版本并且配置了Spark 2.4.8来处理Delta Lake数据，请添加如下参数： <ul style="list-style-type: none"> <li>- 参数名: mgc.delta.spark.version</li> <li>- 参数值: 2</li> </ul> </li> </ul> |

表 1-94 指定日期校验参数配置说明

| 区域   | 参数    | 配置说明                                                                                                                                          |
|------|-------|-----------------------------------------------------------------------------------------------------------------------------------------------|
| 基础信息 | 任务名   | 默认创建名称为“组件类型指定日期校验-4位随机数（字母和数字）”，也可以自定义名称。                                                                                                    |
| 任务配置 | 表组    | 选择需要校验的表所在表组。                                                                                                                                 |
|      | 执行机连接 | <p>Hive、Delta Lake和Hudi组件配置参数。</p> <ul style="list-style-type: none"> <li>• 创建源端统计任务时，选择源端执行机连接。</li> <li>• 创建目的端统计任务时，选择目的端执行机连接。</li> </ul> |

| 区域 | 参数              | 配置说明                                                                                                                             |
|----|-----------------|----------------------------------------------------------------------------------------------------------------------------------|
|    | DLI连接           | 数据湖探索 ( DLI ) 组件配置参数，创建目的端统计任务。选择 <a href="#">创建的DLI连接</a> 。                                                                     |
|    | 元数据连接<br>( 可选 ) | Hive、Delta Lake组件配置参数。<br>创建Hive目的端统计任务时可配置，选择Hive目的端元数据连接。用于判断目的端分区是否存在。<br>创建Delta Lake统计任务时，可以选择与表组所属集群相关的源端元数据连接，用于提升任务处理效率。 |
|    | 执行时间            | 设置作业自动执行时间。当作业激活后，每天都会在设置的时间自动执行。                                                                                                |

| 区域 | 参数   | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|----|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    | 高级选项 | <ul style="list-style-type: none"> <li> <b>并发连接数</b>：进行数据统计时，单个执行机的任务并发数。默认为3，取值范围：1~10。<br/> <b>注意</b><br/>                     在创建阿里云Hive校验任务时，请根据源端数据量和EMR Master节点的规格，合理设置并发连接数。建议按照以下两个条件进行设置：                     <ul style="list-style-type: none"> <li>条件一：同时运行的任务并发数之和不超过元数据节点总核数的70%。</li> <li>条件二：任务分配的总资源不超过执行队列资源。计算如下：<br/>                             内存：num-executors x executor-memory x 并发数<br/>                             核数：num-executors x executor-cores x 并发数<br/>                             例如，源端总数据量为500 G，且包含10000张表，其中8张为50 G、10万分区的大表。EMR Master节点规格为8U32G。                             <ul style="list-style-type: none"> <li>根据条件一：最大并发数 = <math>0.7 \times 8 = 5.6</math>，向下取整得到最大并发数为5。</li> <li>根据条件二：需要在后续执行命令参数配置中，选择<b>Spark-sql</b>，并设置如下参数：                                     <div data-bbox="893 958 1165 1153" style="border: 1px solid #ccc; padding: 5px; margin: 10px 0;"> <p>执行命令</p> <p>baseline    hive    <b>spark-sql</b></p> <p>executor-memory    4G    田</p> <p>master    yarn    田</p> <p>num-executors    20    田</p> <p>executor-cores    2    田</p> <p>driver-memory    10G    田 <a href="#">添加参数</a></p> <p>脚本预览</p> <p>spark-sql --executor-memory 4G --master yarn --num-executors 20 --executor-cores 2 --driver-memory 10G -f</p> </div> <pre>                                         executor-memory = 4G                                         master = yarn                                         num-executors = 20                                         executor-cores = 2                                         driver-memory = 10G                                     </pre> </li> </ul> </li> </ul> </li> <li> <b>单文件SQL条数</b>：每个需要校验的表会产生一条SQL用以获取数据摘要，每批提交给集群处理的数据文件中所包含的SQL条数。默认为10，推荐取值范围：1~50。                 </li> <li> <b>超时时间</b>：每个SQL语句在大数据集群上执行时的最大允许时长，单位为秒（s）。默认为600，推荐取值范围：600~ 7200。                 </li> <li> <b>发送SMN通知</b>：选择是否使用SMN消息通知功能，该功能可以通过电子邮件、短信或者自定义URL等方式及时通知您任务状态。<br/> <b>须知</b> <ul style="list-style-type: none"> <li>使用该功能前，需要前往SMN控制台创建主题。方法请参考<a href="#">创建SMN主题</a>。</li> <li>使用该功能可能产生少量费用，具体费用由SMN服务结算。详情请查看<a href="#">SMN计费说明</a>。</li> </ul> </li> </ul> |

| 区域                               | 参数       | 配置说明                                                                                                                                                                                                                                                      |
|----------------------------------|----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 数据筛选                             | 指定日期     | 选择需要对哪些时间段内的增量数据进行校验。默认统计24小时增量数据。T代表执行时间，T-n代表执行时间前的n x 24小时。<br>勾选 <b>统计连续日期</b> ，可以选择多个统计时间。                                                                                                                                                           |
|                                  | 非分区表处理策略 | 选择对于非分区表的对数处理策略。 <ul style="list-style-type: none"> <li>● <b>全部校验</b>：全量处理所有非分区表，不做筛选。</li> <li>● <b>全部忽略</b>：忽略所有非分区表，不做校验。</li> <li>● <b>按更新时间过滤</b>：只处理更新时间在校验范围内的非分区表。<b>注意</b>：非分区表非insert写入业务，元数据更新时间存在偏差，按更新时间过滤分区时可能遗漏分区，慎重选择该选项！</li> </ul>      |
|                                  | 高级选项     | <b>单SQL字段上限</b> ：每条SQL查询的表的字段上限。单SQL查询的字段数越多，执行时间越长，如果设置的字段数上限过少，则可能形成更多SQL，影响获取校验结果的效率。默认是0，即无上限，取值范围100~500。                                                                                                                                            |
| OBS桶配置检查                         | -        | <ul style="list-style-type: none"> <li>● 如果需要将任务日志文件和内容校验结果上传到OBS桶进行管理和分析，请参考<a href="#">配置OBS桶</a>配置日志桶。配置了日志桶后，会默认使用日志上传功能，将日志数据和内容校验结果上传到指定的OBS桶中。</li> <li>● 如果不需要将任务日志文件和内容校验结果上传到OBS桶，请勾选“<b>我确认仅需在迁移中心Agent侧查看日志和内容校验结果，无需上传至OBS桶</b>”。</li> </ul> |
| 执行脚本<br>(Hive配置参数)               | 安全配置     | 如果大数据集群启用了安全配置（例如Kerberos安全认证），请勾选此选项，并配置安全认证命令。配置前，需要您手动将包含认证密钥的 <b>keytab</b> 文件上传至执行机。 <ul style="list-style-type: none"> <li>● <b>keytab文件地址</b>：输入<b>keytab</b>文件在执行机的存放地址。</li> <li>● <b>keytab文件用户名</b>：输入<b>keytab</b>文件对应的用户名。</li> </ul>        |
|                                  | 执行命令     | 可以通过设置Beeline或Spark-sql命令行参数，对数据进行处理。 <ul style="list-style-type: none"> <li>● <b>Beeline</b>：用于与Hive交互的命令行工具。</li> <li>● <b>Spark-sql</b>：用于执行SQL查询和分析数据。</li> </ul>                                                                                     |
| MaxCompute参数<br>(MaxCompute配置参数) | -        | 根据实际需求，添加MaxCompute参数。参数详情请参考 <a href="#">MaxCompute文档</a> 。                                                                                                                                                                                              |

| 区域                                      | 参数                                              | 配置说明                                                                                                                                                                                                                                                                            |
|-----------------------------------------|-------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 统计配置<br>( DLI配置参数 )                     | 参数                                              | 根据实际需求进行配置，支持的自定义参数详情请参考 <a href="#">自定义参数说明</a> 。                                                                                                                                                                                                                              |
| 命令参数<br>( Delta Lake<br>和Hudi配置<br>参数 ) | 安全配置 ( 仅<br>创建Delta<br>Lake目的端任<br>务时可以配<br>置 ) | <p>如果大数据集群启用了安全配置 ( 例如Kerberos安全认证 )，请勾选此选项，并配置安全认证命令。配置前，需要您手动将包含认证密钥的<code>.keytab</code>文件上传至执行机。</p> <ul style="list-style-type: none"> <li>• <b>keytab文件地址</b>：输入<code>.keytab</code>文件在执行机的存放地址。</li> <li>• <b>keytab文件用户名</b>：输入<code>.keytab</code>文件对应的用户名。</li> </ul> |
|                                         | spark-sql命令                                     | Spark SQL是Spark中用于结构化数据处理的模块。在Spark应用中，可以无缝地使用SQL语句亦或是DataSet API对结构化数据进行查询。更多介绍请参见 <a href="#">SparkSQL原理</a> 。建议保持默认。                                                                                                                                                         |

| 区域 | 参数              | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
|----|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    | spark-submit 命令 | <p>通过Spark基本shell命令，提交Spark应用。命令为：</p> <pre>./bin/spark-submit \ --class &lt;main-class&gt; \ --master &lt;master-url&gt; \ ... # other options &lt;application-jar&gt; \ [application-arguments]</pre> <p>参数解释：</p> <ul style="list-style-type: none"> <li>● --class: Spark应用的类名。</li> <li>● --master: Spark用于所连接的master，如yarn-client, yarn-cluster等。</li> <li>● application-jar: Spark应用的jar包的路径。</li> <li>● application-arguments: 提交Spark应用的所需要的参数（可以为空）。</li> </ul> <p>以下为特定场景下需要添加的参数配置：</p> <ul style="list-style-type: none"> <li>● 如果创建的是阿里云 EMR且涉及Spark 3的Delta Lake源端统计任务，请添加如下参数： <ul style="list-style-type: none"> <li>- 参数名: jars</li> <li>- 参数值: '/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-core_2.12-*.jar,/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-storage-*.jar'</li> </ul> <p><b>注意</b><br/>参数值需要根据实际环境目录和Delta Lake 的版本进行替换。</p> </li> <li>● 如果源端是阿里云 EMR Delta 2.1.0版本并且配置了Spark 2.4.8来处理Delta Lake数据，请添加如下参数： <ul style="list-style-type: none"> <li>- 参数名: mgc.delta.spark.version</li> <li>- 参数值: 2</li> </ul> </li> </ul> |

表 1-95 指定时间校验参数配置说明

| 区域   | 参数      | 配置说明                                                                                                                                                              |
|------|---------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 基础信息 | 任务名     | 默认创建名称为“组件类型指定时间校验-4位随机数（字母和数字）”，也可以自定义名称。                                                                                                                        |
| 任务配置 | 表组      | 选择需要校验的表所在表组。                                                                                                                                                     |
|      | HBase连接 | HBase组件和CloudTable(HBase)组件配置参数。 <ul style="list-style-type: none"> <li>● 创建源端统计任务时，选择源端HBase连接。</li> <li>● 创建目的端统计任务时，选择目的端HBase或CloudTable(HBase)连接。</li> </ul> |

| 区域                  | 参数   | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|---------------------|------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                     | 高级选项 | <ul style="list-style-type: none"> <li>● <b>并发连接数</b>：进行数据统计时，单个执行机的任务并发数。默认为3，取值范围：1~10。</li> <li>● <b>超时时间</b>：每个SQL语句在大数据集群上执行时的最大允许时长，单位为秒（s）。默认为600，推荐取值范围：600~ 7200。</li> <li>● <b>发送SMN通知</b>：选择是否使用SMN消息通知功能，该功能可以通过电子邮件、短信或者自定义URL等方式及时通知您任务状态。<br/><b>须知</b> <ul style="list-style-type: none"> <li>- 使用该功能前，需要前往SMN控制台创建主题。方法请参考<a href="#">创建SMN主题</a>。</li> <li>- 使用该功能可能产生少量费用，具体费用由SMN服务结算。详情请查看<a href="#">SMN计费说明</a>。</li> </ul> </li> </ul> |
| 数据筛选                | 统计范围 | 选择需要对哪段时间内的数据进行校验的。                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| OBS桶配置检查            | -    | <ul style="list-style-type: none"> <li>● 如果需要将任务日志文件和内容校验结果上传到OBS桶进行管理和分析，请参考<a href="#">配置OBS桶</a>配置日志桶。配置了日志桶后，会默认使用日志上传功能，将日志数据和内容校验结果上传到指定的OBS桶中。</li> <li>● 如果不需要将任务日志文件和内容校验结果上传到OBS桶，请勾选“<b>我确认仅需在迁移中心Agent侧查看日志和内容校验结果，无需上传至OBS桶</b>”。</li> </ul>                                                                                                                                                                                                      |
| 统计配置<br>(HBase配置参数) | 运行模式 | 支持以下两种运行模式： <ul style="list-style-type: none"> <li>● <b>YARN 模式</b>：适用于大型分布式环境，可以充分利用集群资源，提高任务的并行度和效率。<br/><b>说明</b><br/>MRS 3.1.0版本普通集群使用YARN模式时，需要进行换包操作。操作方法请参考<a href="#">使用YARN模式换包操作指南</a>。</li> <li>● <b>Local 模式</b>：适用于小规模数据集或开发测试环境，可以快速进行调试和验证。</li> </ul>                                                                                                                                                                                          |
|                     | 参数   | 根据所选的运行模式和实际需求，添加命令行参数。                                                                                                                                                                                                                                                                                                                                                                                                                                        |

**步骤7** 配置完成后，单击“**保存**”按钮，完成统计任务的创建。创建成功后，系统会自动将参数配置同步到MgC Agent。您可以在**任务列表**查看创建的统计任务及其配置同步状态。

**步骤8** 配置同步完成后，您可以选择以下两种方式执行统计任务：

- **自动定时执行**：按照设置的执行时间，自动、定时执行作业。
  - a. 在任务列表，单击统计任务**调度状态**列的“**激活**”按钮，弹出**激活**提示窗。
  - b. 单击“**确认**”按钮，激活自动化执行作业。
- **手动执行**：手动立即触发一次作业的执行。



| 统计指标                      | 说明                                         |
|---------------------------|--------------------------------------------|
| Count已达标                  | 统计当天在数据校验中，Count（计数）达标的表的数量。               |
| Count未达标                  | 统计当天在数据校验中，Count未达标的表的数量。                  |
| Count已达标，Sum已达标           | 统计当天在数据校验中，Count和Sum（求和）都达标的表的数量。          |
| Count已达标，Sum未达标           | 统计当天在数据校验中，Count达标但Sum未达标的表的数量。            |
| Count未达标，Sum未达标           | 统计当天在数据校验中，Count和Sum都未达标的表的数量。             |
| Count已达标，Sum已达标，AllSum已达标 | 统计当天在数据校验中，Count、Sum和AllSum（全部求和）都达标的表的数量。 |
| Count已达标，Sum已达标，AllSum未达标 | 统计当天在数据校验中，Count和Sum达标但AllSum未达标的表的数量。     |
| Count已达标，Sum未达标，AllSum未达标 | 统计当天在数据校验中，Count达标但Sum和AllSum未达标的表的数量。     |
| Count未达标，Sum未达标，AllSum未达标 | 统计当天在数据校验中，Count、Sum和AllSum都未达标的表的数量。      |

## 1.12.7 查看并导出校验结果

当源端与目的端的统计任务都执行完成后，在**校验结果**页面可以查看并导出任务的执行结果。校验结果将按照任务开始执行的日期进行分批，如果单日多次执行校验任务，系统只会保留最后一次的校验结果，之前的执行结果会被后续的执行所覆盖。

### 前提条件

已**创建统计任务并执行校验**。

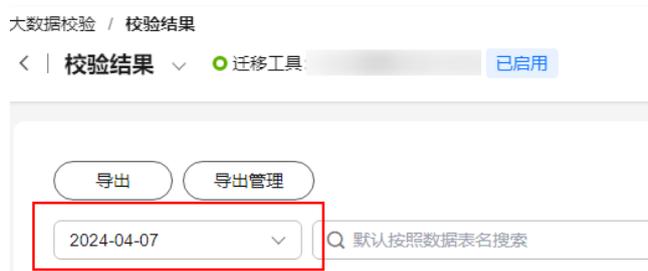
### 使用须知

当执行求和操作（如sum或allsum）时，如果结果值的位数超过了Double类型的有效位数，会导致精度丢失。并且在处理非常大的数值时，由于浮点数的精度限制，可能会导致实际存在的差异因为超出了Double类型能表示的有效位数而无法被感知到，从而使得差异率计算结果为0。

### 查看校验结果

- 步骤1** 登录**迁移中心管理控制台**。在左侧导航栏的项目下拉列表中选择创建的**大数据迁移项目**。
- 步骤2** 在左侧导航栏选择“**实施>大数据校验**”，进入**大数据校验**页面。
- 步骤3** 在**功能入口**区域，单击“**校验结果**”，进入**校验结果**页面。

**步骤4** 通过执行日期，筛选出要查看的执行批次。



**步骤5** 在校验结果列表中，可以进行如下操作：

- 查看校验详情：单击分区达标详情列的“查看”按钮，展开该数据表的作业详情。单击操作列的“查看详情”，进入校验详情页面，查看该数据表的校验详细信息。
- 查看执行历史及状态：单击操作列的“查看历史”，可以查看该表的所有执行批次和达标状态。

----结束

## 导出校验结果

**步骤1** 在校验结果页面，单击“导出 > 导出校验结果”按钮，弹出导出校验结果窗口。

**步骤2** 选择需要导出的批次、表组名以及类型，单击“确认”按钮，系统开始导出校验结果。

**步骤3** 导出完成后，单击“导出管理”按钮，右侧弹出导出管理窗口。

**步骤4** 单击操作列的“下载”，可以将对应表的校验结果下载到本地。

----结束

## 导出数据分析报告

**步骤1** 在校验结果页面，单击“导出 > 导出数据分析”按钮，弹出导出数据分析报告窗口。

图 1-65 导出数据分析报告



**步骤2** 选择需要导出日期范围、导出内容，单击“确定”按钮，系统开始导出数据分析报告结果。

图 1-66 选择导出的数据报告范围



**步骤3** 导出完成后，单击“导出管理”按钮，右侧弹出导出管理窗口。

**步骤4** 单击操作列的“下载”，可以将对应表的校验结果下载到本地。

----结束

## 1.12.8 自定义参数说明

在创建数据湖探索 DLI校验任务时，支持的自定义参数及说明参见表1-97。

表 1-97 DLI 校验任务支持的自定义参数说明

| 参数名称                                 | 默认值       | 描述                                                                                                                                                                                    |
|--------------------------------------|-----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| mgc.mc2dli.table.partition.enable    | true      | DLI表分区为空或者分区不存在时需要通过查询DLI的元数据去判断分区存不存在。<br><ul style="list-style-type: none"> <li>• true: 查询DLI表的分区，空分区校验状态为成功，分区不存在状态为失败。</li> <li>• false: 不查询DLI表的分区，空分区或者分区不存在校验状态都为成功</li> </ul> |
| spark.sql.files.maxRecordsPerFile    | 0         | 要写入单个文件的最大记录数。如果该值为零或为负，则没有限制。                                                                                                                                                        |
| spark.sql.autoBroadcastJoinThreshold | 209715200 | 配置执行连接时显示所有工作节点的表的最大字节大小。通过将此值设置为“-1”，可以禁用显示。<br><b>说明</b><br>当前仅支持运行命令ANALYZE TABLE COMPUTE statistics noscan的配置单元元存储表，和直接根据数据文件计算统计信息的基于文件的数据源表。                                    |
| spark.sql.shuffle.partitions         | 200       | 为连接或聚合过滤数据时使用的默认分区数。                                                                                                                                                                  |

| 参数名称                                             | 默认值       | 描述                                                                                                                                                                                                                                                                                                                                                                                                                           |
|--------------------------------------------------|-----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| spark.sql.dynamicPartitionOverwrite.enabled      | false     | <p>当前配置设置为“false”时，DLI在覆盖写之前，会删除所有符合条件的分区。例如，分区表中有一个“2021-01”的分区，当使用INSERT OVERWRITE语句向表中写入“2021-02”这个分区的数据时，会把“2021-01”的分区数据也覆盖掉。</p> <p>当前配置设置为“true”时，DLI不会提前删除分区，而是在运行时覆盖那些有数据写入的分区。</p>                                                                                                                                                                                                                                 |
| spark.sql.files.maxPartitionBytes                | 134217728 | 读取文件时要打包到单个分区中的最大字节数。                                                                                                                                                                                                                                                                                                                                                                                                        |
| spark.sql.badRecordsPath                         | -         | Bad Records的路径。                                                                                                                                                                                                                                                                                                                                                                                                              |
| spark.sql.legacy.correlated.scalar.query.enabled | false     | <ul style="list-style-type: none"> <li>● 该参数设置为true:                             <ul style="list-style-type: none"> <li>- 当子查询中数据不重复的情况下，执行关联子查询，不需要对子查询的结果去重。</li> <li>- 当子查询中数据重复的情况下，执行关联子查询，会提示异常，必须对子查询的结果做去重处理，比如max(),min()。</li> </ul> </li> <li>● 该参数设置为false:                             <ul style="list-style-type: none"> <li>- 不管子查询中数据重复与否，执行关联子查询时，都需要对子查询的结果去重，比如max(),min()，否则提示异常。</li> </ul> </li> </ul> |

# 2 旧版

## 2.1 权限管理

### 2.1.1 创建用户并授权使用 MgC

如果您需要对您所拥有的MgC进行精细的权限管理，您可以使用[统一身份认证服务](#)（Identity and Access Management，简称IAM），通过IAM，您可以：

- 根据企业的业务组织，在您的华为云账号中，给企业中不同职能部门的员工创建IAM用户，让员工拥有唯一安全凭证，并使用MgC资源。
- 根据企业用户的职能，设置不同的访问权限，以达到用户之间的权限隔离。
- 将MgC资源委托给更专业、高效的其他华为云账号或者云服务，这些账号或者云服务可以根据权限进行代运维。

如果华为云账号已经能满足您的要求，不需要创建独立的IAM用户，您可以跳过本章节，不影响您使用MgC服务的其它功能。

本章节为您介绍对用户授权的方法，操作流程如[示例流程](#)所示。

#### 前提条件

给用户组授权之前，请您了解用户组可以添加的MgC系统策略，并结合实际需求进行选择。

MgC支持的系统策略及策略间的对比，请参见：[MgC系统策略](#)。若您需要对除MgC之外的其它服务授权，IAM支持服务的所有策略请参见[权限策略](#)。

## 示例流程

图 2-1 给用户授权 MgC 权限流程



### 1. 创建用户组并授权

- **系统策略**: 在IAM控制台创建用户组，并根据**MgC系统策略说明**以及实际的权限要求，为用户组授权MgC系统策略，授权范围方案选择“所有资源”。

表 2-1 MgC 系统策略说明

| 策略名称                | 描述                                       | 策略类别 | 策略内容                                    |
|---------------------|------------------------------------------|------|-----------------------------------------|
| MgC FullAccess      | 迁移中心管理员权限，拥有操作MgC的所有权限。                  | 系统策略 | <a href="#">MgC FullAccess策略内容</a>      |
| MgC ReadOnlyAccess  | 迁移中心只读权限，仅能查看MgC资源，无法进行操作。               | 系统策略 | <a href="#">MgC ReadOnlyAccess策略内容</a>  |
| MgC DiscoveryAccess | 迁移中心资源发现操作权限，拥有操作资源发现功能的权限和只读权限。         | 系统策略 | <a href="#">MgC DiscoveryAccess策略内容</a> |
| MgC AssessAccess    | 迁移中心评估操作权限，拥有操作评估功能、资源发现功能的权限和只读权限。      | 系统策略 | <a href="#">MgC AssessAccess策略内容</a>    |
| MgC MigrateAccess   | 迁移中心迁移操作权限，拥有操作迁移功能、评估功能、资源发现功能的权限和只读权限。 | 系统策略 | <a href="#">MgC MigrateAccess策略内容</a>   |

| 策略名称                    | 描述                                      | 策略类别 | 策略内容                                        |
|-------------------------|-----------------------------------------|------|---------------------------------------------|
| MgC AppDiscovery Access | 迁移中心资源采集操作权限，拥有操作资源采集功能、资源发现功能的权限和只读权限。 | 系统策略 | <a href="#">MgC AppDiscovery Access策略内容</a> |
| MgC MrrAccess           | 迁移中心业务验证操作权限，拥有业务验证功能的权限和只读权限。          | 系统策略 | <a href="#">MgC MrrAccess策略内容</a>           |

- **自定义策略：**如果IAM用户只需要拥有迁移中心 MgC部分操作权限，则使用自定义策略，参见[MgC自定义策略](#)。

## 2. 创建用户并加入用户组

在IAM控制台创建用户，并将其加入[1.创建用户组并授权](#)中创建的用户组。

### 须知

当IAM用户首次使用MgC的特定功能时，需要通过以下两种方式授予委托权限：

- **推荐方式：**请管理员登录MgC控制台，进入相关功能界面，同意相关委托权限。
- **备选方式：**请管理员登录“统一身份认证服务 IAM”服务，为当前IAM用户配置创建委托所需的权限，确保其拥有MgC相关系统策略以及"iam:agencies:createAgency"、"iam:permissions:grantRoleToAgency"、"iam:roles:createRole"和"iam:roles:updateRole"的权限。

## 3. 用户登录并验证权限

新创建的用户登录控制台，切换至授权区域，验证权限：

- 在“**服务列表**”中选择迁移中心 MgC，进入MgC主界面，根据您授予的权限可以进行对应操作，表示授予的权限已生效。
- 在“**服务列表**”中选择除迁移中心 MgC外的任一服务，若提示权限不足，表示授予的权限已生效。

## MgC FullAccess 策略内容

```
{
 "Version": "1.1",
 "Statement": [
 {
 "Action": [
 "mgc:*",
 "iam:agencies:listAgencies",
 "iam:roles:listRoles",
 "iam:quotas:listQuotas",
 "iam:permissions:listRolesForAgency"
],
 "Effect": "Allow"
 }
]
}
```

## MgC ReadOnlyAccess 策略内容

```
{
 "Version": "1.1",
```

```
"Statement": [
 {
 "Effect": "Allow",
 "Action": [
 "mgc:*:query*"
]
 }
]
```

## MgC DiscoveryAccess 策略内容

```
{
 "Version": "1.1",
 "Statement": [
 {
 "Effect": "Allow",
 "Action": [
 "mgc:*:query*",
 "mgc:*:discovery"
]
 }
]
}
```

## MgC AssessAccess 策略内容

```
{
 "Version": "1.1",
 "Statement": [
 {
 "Action": [
 "mgc:*:query*",
 "mgc:*:discovery",
 "mgc:*:assess",
 "iam:agencies:listAgencies",
 "iam:roles:listRoles",
 "iam:quotas:listQuotas",
 "iam:permissions:listRolesForAgency"
],
 "Effect": "Allow"
 }
]
}
```

## MgC MigrateAccess 策略内容

```
{
 "Version": "1.1",
 "Statement": [
 {
 "Action": [
 "mgc:*:query*",
 "mgc:*:discovery",
 "mgc:*:assess",
 "mgc:*:migrate",
 "iam:agencies:listAgencies",
 "iam:roles:listRoles",
 "iam:quotas:listQuotas",
 "iam:permissions:listRolesForAgency"
],
 "Effect": "Allow"
 }
]
}
```

## MgC AppDiscoveryAccess 策略内容

```
{
 "Version": "1.1",
 "Statement": [
 {
 "Effect": "Allow",
 "Action": [
 "mgc:*:query*",
 "mgc:*:discovery",
 "mgc:*:appdiscovery"
]
 }
]
}
```

## MgC MrrAccess 策略内容

```
{
 "Version": "1.1",
 "Statement": [
 {
 "Action": [
 "mgc:*:query*",
 "mgc:mrr:query",
 "mgc:mrr:update",
 "mgc:mrr:export",
 "mgc:mrr:import",
 "mgc:mrr:upgrade",
 "mgc:mrr:delete",
 "mgc:mrr:check"
],
 "Effect": "Allow"
 }
]
}
```

### 2.1.2 MgC 自定义策略

如果系统预置的MgC权限，不满足您的授权要求，可以创建自定义策略。

目前华为云支持以下两种方式创建自定义策略：

- 可视化视图创建自定义策略：无需了解策略语法，按可视化视图导航栏选择云服务、操作、资源、条件等策略内容，可自动生成策略。
- JSON视图创建自定义策略：可以在选择策略模板后，根据具体需求编辑策略内容；也可以直接在编辑框内编写JSON格式的策略内容。

具体创建步骤请参见：[创建自定义策略](#)。本章为您介绍常用的MgC自定义策略样例。

#### MgC 自定义策略样例

- 平台采集、资源管理

```
{
 "Version": "1.1",
 "Statement": [
 {
 "Effect": "Allow",
 "Action": [
 "mgc:*:query*",
 "mgc:*:discovery"
]
 }
]
}
```

- ```

    }
  ]
}

```
- TCO评估、主机评估


```

{
  "Version": "1.1",
  "Statement": [
    {
      "Action": [
        "mgc*:query*",
        "mgc*:discovery",
        "mgc*:assess",
        "iam:agencies:listAgencies",
      ],
      "Effect": "Allow"
    }
  ]
}

```
 - 应用关联分析


```

{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "mgc*:query*",
        "mgc*:discovery",
        "mgc*:appdiscovery"
      ]
    }
  ]
}

```
 - 迁移 workflow


```

{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "mgc*:query*",
        "mgc*:discovery",
        "mgc*:assess",
        "mgc*:migrate",
        "iam:agencies:listAgencies"
      ]
    }
  ]
}

```

2.1.3 委托权限说明

概述

当您启用特定功能时，需要您确认并同意授予必要的权限给MgC，以便我们能够为您提供完整的服务。本节为您介绍需要进行权限委托的场景以及自定义策略详情。

委托授权策略说明：

- **创建自定义策略**：如果您尚未设定任何自定义的权限委托策略，系统会自动为您创建一个标准的自定义策略。
- **更新自定义策略**：对于已经存在的自定义策略，系统将进行完整性检查。若发现策略内容不全面或缺少必要元素，系统将自动进行补充。

跨可用区迁移

场景	委托对象	自定义策略名称	细粒度最小使用权限
跨可用区迁移	MgC	MgC AzMigrationAgencyPolicy	ecs:cloudServers:showServer (查询云服务器详情) ecs:flavors:get (查询云服务器规格) ecs:cloudServerFlavors:get (查询云服务器规格详情和扩展信息列表) ecs:cloudServerQuotas:get (查询租户配额) ecs:servers:list (查询云服务器列表) ecs:cloudServers:list (查询云服务器详情列表) ecs:servers:stop (关闭云服务器) ecs:cloudServers:listServerInterfaces (查询云服务器网卡信息) ecs:cloudServers:createServers (创建云服务器) ecs:cloudServers:listServerBlockDevices (查询弹性云服务器磁盘信息) ecs:cloudServerNics:update (云服务器网卡配置私有IP) ecs:availabilityZones:list (查询可用区列表) ecs:servers:start (启动云服务器) ecs:cloudServers:changeNetworkInterface (更新云服务器指定网卡属性) ecs:serverInterfaces:get (查询云服务器网卡) ecs:cloudServers:get (查询云服务器详情) vpc:publicIps:create (创建弹性公网IP) vpc:publicIps:update (更新弹性公网IP) vpc:subnets:get (查询子网列表或详情) vpc:networks:get (查询网络列表或详情) vpc:publicIps:list (查询弹性公网IP) vpc:publicIps:get (查询弹性公网IP详情) vpc:ports:get (查询端口列表或详情) vpc:ports:delete (删除端口)

场景	委托对象	自定义策略名称	细粒度最小使用权限
			vpc:ports:update (更新端口) vpc:ports:create (创建端口) evs:types:get (查询云硬盘类型) evs:volumes:list (查询云硬盘列表) cbr:vaults:get (查询指定存储) cbr:vaults:list (查询存储库列表) cbr:vaults:create (创建存储库) cbr:vaults:addResources (添加资源) cbr:vaults:backup (执行备份) cbr:backups:list (查询备份列表) cbr:tasks:list (查询任务列表) cbr:tasks:get (查询单个任务) cbr:backups:delete (删除备份) cbr:backups:get (查询指定备份) cbr:vaults:delete (删除存储库) ims:wholeImages:create (制作整机镜像) ims:images:list (查询镜像列表) ims:images:delete (删除镜像) ims:images:get (查询镜像详情) ims:serverImages:create (制作镜像)

上云成本分析

场景	委托对象	自定义策略名称	细粒度最小使用权限
创建上云成本分析任务	MgC	MgC TcoAgencyPolicy	ecs:cloudServerFlavors:get (查询云服务器规格详情和扩展信息列表) evs:types:get (查询云硬盘类型) ims:*.get* (ims查询权限) ims:*.list* (ims查询权限)

规格评估

场景	委托对象	自定义策略名称	细粒度最小使用权限
规格评估	MgC	MgC ServerAssessAge ncyPolicy	ecs:cloudServerFlavors:get (查询云服务器规格详情和扩展信息列表) ims:images:list (查询镜像列表) evs:types:get (查询云硬盘类型) deh:dedicatedHosts:get (获取专属主机详情) deh:dedicatedHosts:list (查询专属主机列表)

主机绑定目的端

场景	委托对象	自定义策略名称	细粒度最小使用权限
主机绑定目的端	MgC	MgC ServerBindTarget AgencyPolicy	ecs:cloudServers:showServer (查询云服务器详情) evs:volumes:list (查询云硬盘列表) ecs:cloudServerFlavors:get (查询云服务器规格详情和扩展信息列表)

主机迁移 workflow

场景	委托对象	自定义策略名称	细粒度最小使用权限
主机迁移 workflow	MgC	MgC ServerMigrationAgencyPolicy	ecs:cloudServers:showServer (查询云服务器详情) ecs:cloudServers:createServers (创建云服务器) sms:server:migrationServer (迁移源端服务器) sms:server:queryServer (查看源端服务器) ecs:cloudServers:list (查看云服务器列表) ecs:cloudServers:listServerBlockDevices (查询弹性云服务器磁盘信息) ecs:cloudServerQuotas:get (查询租户配额) vpc:publicIps:create (创建弹性公网IP) ecs:cloudServers:get (查询云服务器详情) ecs:cloudServers:changeVpc (云服务器切换虚拟私有网络) ecs:cloudServers:attach (弹性云服务器云主机挂载磁盘) ecs:cloudServers:start (批量启动云服务器) ecs:cloudServers:detachVolume (卸载指定弹性云服务器的磁盘) ecs:cloudServers:stop (批量关闭云服务器) ecs:servers:unlock (解锁定云服务器) evs:volumes:delete (删除云硬盘) evs:volumes:use (云硬盘挂载和卸载依赖) evs:volumes:get (查询单个云硬盘详情) vpc:privateIps:list (查询私有IP列表) ims:images:get (查询镜像详情)

购买资源

场景	委托对象	自定义策略名称	细粒度最小使用权限
购买资源	MgC	MgC PurchaseAgencyPolicy	eps:resources:add (企业项目资源迁入) ecs:cloudServers:createServers (创建云服务器) evs:volumes:list (查询云硬盘列表) ecs:cloudServerFlavors:get (查询云服务器规格详情和扩展信息列表) ecs:cloudServers:list (查询云服务器详情列表) vpc:publicIps:update (更新弹性公网 IP) vpc:publicIps:create (创建弹性公网 IP)

配置主机购买模板

场景	委托对象	自定义策略名称	细粒度最小使用权限
配置主机购买模板	MgC	MgC PurchaseTemplateAgencyPolicy	iam:projects:listProjects (查询租户项目) eps:enterpriseProjects:list (查看企业项目列表) vpc:subnets:get (查询子网列表或详情) vpc:securityGroups:get (查询安全组列表或详情) vpc:vpcs:get (查询虚拟私有云详情)

迁移方案配置

场景	委托对象	自定义策略名称	细粒度最小使用权限
主机标准迁移模板 - 导入OBS桶里的目的端配置	MgC	MgC ImportTargetConfigurationAgencyPolicy	obs:object:GetObject (获取对象内容、获取对象元数据) obs:bucket:ListBucket (列举桶内对象) obs:bucket:ListAllMyBuckets (获取桶列表)

场景	委托对象	自定义策略名称	细粒度最小使用权限
主机标准迁移模板 - 导出支持的目的地配置		MgC ExportTargetConfigurationAgencyPolicy	ims:images:list (查询镜像列表) ecs:cloudServerFlavors:get (查询云服务器规格详情和扩展信息列表)
对象存储批量迁移模板-配置目的地		MgC ListObsBucketsAgencyPolicy	obs:bucket:ListBucket (列举桶内对象) obs:bucket:ListAllMyBuckets (获取桶列表)

创建迁移集群

场景	委托对象	自定义策略名称	细粒度最小使用权限
创建迁移集群	OMS	OMS ObsMigrationAgencyPolicy	ecs:cloudServers:createServers (创建云服务器) ecs:cloudServers:listServerInterfaces (查询云服务器网卡信息) ecs:cloudServers:showServer (查询云服务器详情) ecs:cloudServers:deleteServers (删除云服务器) ecs:cloudServers:list (查询云服务器详情列表) nat:natGateways:create (创建NAT网关) nat:natGateways:get (查询NAT网关详情) nat:natGateways:delete (删除NAT网关) nat:snatRules:create (创建SNAT规则) nat:snatRules:get (查询SNAT规则详情) nat:dnatRules:list (查询DNAT规则列表) nat:snatRules:list (查询SNAT规则列表) nat:snatRules:delete (删除SNAT规则) nat:natGateways:list (查询NAT网关列表) vpc:securityGroups:create (创建安全组) vpc:securityGroups:delete (删除安全组) vpc:securityGroups:get (查询安全组列表或详情) vpc:securityGroupRules:create (创建安全组规则) vpc:securityGroupRules:get (查询安全组规则列表或详情) vpc:securityGroupRules:delete (删除安全组规则) vpcep:epservices:create (创建终端节点服务)

场景	委托对象	自定义策略名称	细粒度最小使用权限
			vpcep:epservices:get (查询终端节点服务详情) vpcep:permissions:list (查询终端节点服务的白名单列表) vpcep:connections:list (查询连接终端节点服务的连接列表) vpcep:epservices:list (查询终端节点服务列表) vpcep:epservices:delete (删除终端节点服务) vpcep:connections:update (接受或拒绝终端节点的连接) vpcep:permissions:update (批量添加或移除终端节点服务的白名单) lts:topics:create (创建日志主题) lts:topics:delete (删除指定日志主题) lts:groups:create (创建日志组) lts:groups:get (查询指定日志组) lts:groups:delete (删除指定日志组)
	ECS	ECS ObsMigrationAgencyPolicy	apm:icmgr:* (apm采集组件的所有权限)

存储迁移 workflow

场景	委托对象	自定义策略名称	细粒度最小使用权限
存储迁移 workflow	MgC	-	OMS Administrator (系统角色, OMS的admin角色)

RVTools 导入

场景	委托对象	自定义策略名称	细粒度最小使用权限
创建RVTools 导入任务	MgC	MgC OfflineCollectionAge ncyPolicy	obs:object:GetObject (获取对象 内容、获取对象元数据) obs:bucket:ListBucket (列举桶 内对象) obs:bucket:ListAllMyBuckets (获取桶列表)

2.1.4 IAM 用户权限

MgC在统一身份认证服务（IAM）能力基础上，为用户提供细粒度的权限管理功能，帮助用户灵活便捷地对账号下的IAM用户设置不同的MgC资源权限，结合权限策略实现权限隔离。

以下是使用MgC不同功能时，IAM用户需要配置的权限。为IAM用户配置权限的方法请参考[创建自定义策略](#)。

功能	权限Action
配置主机购买模板	vpc:vpcs:list (查询虚拟私有云列表) vpc:subnets:get (查询子网列表或详情) vpc:publicips:list (查询弹性公网IP) vpc:securityGroups:get (查询安全组列表或详情) eps:enterpriseProjects:list (查看企业项目列表) ecs:availabilityZones:list (查询可用区列表)
主机迁移 workflow	vpc:vpcs:list (查询虚拟私有云列表) vpc:vpcs:get (查询虚拟私有云详情) vpc:subnets:get (查询子网列表或详情) vpc:publicips:list (查询弹性公网IP) vpc:publicips:get (查询弹性公网IP详情) vpc:securityGroups:get (查询安全组列表或详情) eps:enterpriseProjects:list (查看企业项目列表) eps:enterpriseProjects:get (查看有企业项目详情)

功能	权限Action
主机规格评估	ecs:cloudServerFlavors:get (查询云服务器规格详情和扩展信息列表) ecs:cloudServers:list (查询云服务器详情列表) ecs:cloudServers:showServer (查询云服务器详情) ecs:flavors:get (查询云服务器规格) ims:images:list (查询镜像列表) ims:images:get (查询镜像详情) evs:volumes:list (查询云硬盘列表) evs:types:get (查询云硬盘类型)
AZ迁移 workflow	ecs:availabilityZones:list (查询可用区列表)
配置TCO对比产品映射	ecs:cloudServerFlavors:get (查询云服务器规格详情和扩展信息列表) ims:images:list (查询镜像列表) evs:types:get (查询云硬盘类型)
存储迁移 workflow	Tenant Guest (全部云服务只读权限) (除IAM权限) OMS Administrator (对象存储迁移服务所有权限)
迁移集群	Tenant Guest (全部云服务只读权限) (除IAM权限) OMS Administrator (对象存储迁移服务所有权限) nat:natGateways:list (查询NAT网关列表) smn:topic:list (主题查询权限)
委托授权	iam:agencies:listAgencies (查询指定条件下的委托列表信息) iam:roles:listRoles (查询权限列表) iam:quotas:listQuotas (查询配额) iam:permissions:listRolesForAgency (查询委托的所有权限) iam:agencies:createAgency (创建委托) iam:permissions:grantRoleToAgency (为委托授予指定权限) iam:roles:createRole (创建自定义策略) iam:roles:updateRole (修改自定义策略)

2.2 配置管理

2.2.1 项目管理

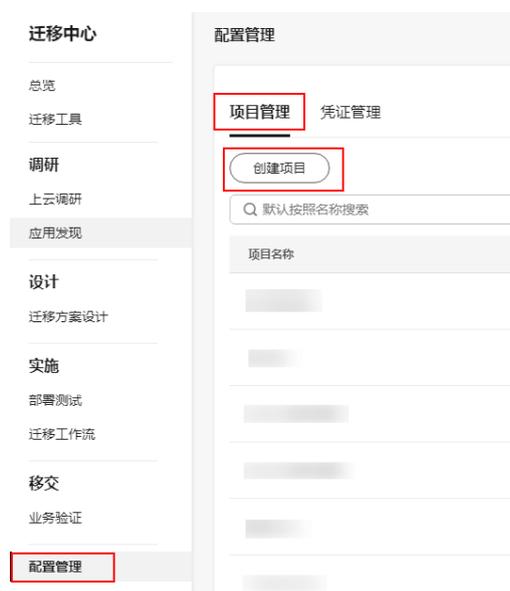
操作场景

通过创建独立的项目，可以将不同的操作场景进行区分和隔离，从而实现更好的组织管理和资源分配。例如：可为阿里云迁移至华为云创建一个新的项目，里面将包含阿里云平台的主机、数据库、组件等信息。

创建项目

步骤1 登录**迁移中心**管理控制台。

步骤2 单击左侧导航栏的“**配置管理**”，进入**配置管理**页面。选择“**项目管理**”页签，单击“**创建项目**”。



步骤3 在弹出的窗口中，自定义**项目名称**，仔细查看并了解项目类型的适用场景后，选择项目类型，单击“**确认**”。创建项目成功，可在项目管理列表查看。



----结束

修改项目

- 步骤1** 单击操作列的“修改”按钮，弹出“修改项目”窗口。
- 步骤2** 可以修改项目名称、输入项目描述、设置项目的开始时间和结束时间，单击“确认”，完成修改。

修改项目

* 项目名称

项目描述 0/255

开始时间 请选择日期时间

结束时间 请选择日期时间

确认 取消

----结束

归档项目

归档后的项目将在当前项目下拉列表中隐藏，无法选择。单击操作列的“归档”按钮，可将项目归档。如想重新选择归档后的项目，单击“取消归档”即可。

项目管理 凭证管理

创建项目

默认按照名称搜索

项目名称	项目描述	创建时间	开始时间/结束时间	操作
	-	2023/10/17 15:46:31 GMT+08:00	-	修改 归档

说明

当前项目不允许归档。

2.2.2 凭证管理

凭证是资源采集的唯一认证方式，本节介绍如何添加用于发现云上资源（公网发现）所需要的凭证。内网发现、手动添加所需的凭证添加方式请参考[在Edge添加凭证](#)。

认证方式

当前仅支持添加公有云资源类型的凭证。支持的认证方式参见[表2-2](#)。

表 2-2 公有云认证方式

资源类型	认证方式	说明
公有云	AK/SK	适用于添加华为云、阿里云、AWS、腾讯云、七牛云和金山云等云平台凭证。各云平台的AK/SK获取方法请参考 如何获取云平台凭证 。
	配置文件	适用于添加谷歌云平台凭证。配置文件需要上传谷歌云 服务账号 凭据，该凭证必须为json格式文件，且大小不能超过4KB。
	ID/密钥	适用于添加Azure平台凭证。添加凭证所需的信息获取方法请参见 如何获取添加Azure凭证所需的信息 。

注意

云上凭证的存储有效期为7天，超过7天后，凭证会过期，需要重新输入。

操作步骤

- 步骤1** 登录[迁移中心](#)管理控制台。
- 步骤2** 单击左侧导航栏的“配置管理”，进入配置管理页面。在页面左上角的当前项目下拉列表中选择[迁移项目](#)。
- 步骤3** 选择“凭证管理”页签，单击“添加凭证”，弹出“添加云平台凭证”窗口。



- 步骤4** 按照窗口提示，选择资源类型后，选择认证方式并输入/上传相应凭证，凭证名称为用户自定义，单击“确认”。

添加完成后，在凭证管理列表可以看到添加的凭证。

添加云平台凭证

* 使用类型

* 类型

* 凭证名称

* 认证方式

* AK

* SK

----结束

2.3 上云调研

2.3.1 上云成本分析（凭证）

通过输入源端云平台凭证，可以自动完成源端云厂商消费账单分析，根据源端产品规格及用量自动匹配华为云上的产品及用量，并对比源端云厂商使用费用和华为云使用费用，为您云间迁移提供参考。

目前支持通过凭证进行上云成本分析的源端云厂商包括：**AWS、阿里云**。

须知

进行上云成本分析时，请确保服务器数量不超过9999台。

创建分析任务

- 步骤1** 登录**迁移中心**管理控制台。
- 步骤2** 单击左侧导航栏“**调研>上云调研**”，进入“**上云调研**”页面。在页面左上角的当前项目下拉列表中选择**迁移项目**。
- 步骤3** 选中**上云成本分析（凭证）**框，单击“**创建分析任务**”按钮，右侧弹出“**创建分析任务**”窗口。



步骤4 根据表2-3，配置参数。

表 2-3 参数说明

参数	说明
任务名称	用户自定义。
云厂商	选择源端云厂商。 说明 当前源端云厂商仅支持AWS和阿里云。
区域	选择源端资源所在区域。
调研方式	<ul style="list-style-type: none"> “AK/SK”方式，需要选择“凭证”、“对比时间段”和“对比产品”，如未提前添加凭证，请单击“新建凭证”，填写相应参数，校验并保存凭证。 说明 <ul style="list-style-type: none"> 新建凭证中的AK/SK输入源端账号的AK/SK。 源端账号拥有只读权限即可。 可能产生接口调用相关费用。 “表格上传”方式，通过填写TCO任务模板的方式上传源端信息进行评估分析。 须知 模板中的主机数量不能超过9999台。
目的端区域	选择目的端所在区域。

步骤5 配置完成后，单击“确认”按钮，提交上云成本分析任务。

提交成功后可以在上云成本分析（凭证）任务列表中看到提交的任务，单击**任务名称**，可以查看任务进度，等待任务运行完成后即可[下载账单](#)、[查看配置](#)、[产品对比](#)，进行后续详细分析。

----**结束**

下载账单

系统会从产品计费项的维度分析您指定时间段内的账单，解析出每款产品的规格、月均使用量以及月均消耗，您可以在上云成本分析任务完成后下载您的账单，在本地进一步分析。

步骤1 在上云成本分析（凭证）任务列表，单击**任务名称**。

步骤2 单击操作列的“**下载账单**”按钮，下载Excel版账单。



----结束

查看配置

系统会根据解析出来的产品规格及用量自动匹配华为云上的产品及用量，在匹配策略上系统坚持在不低于您当前配置的前提下优先考虑成本的原则，给出华为云上的产品规格及用量配置。

您可以通过**查看配置**进入产品映射明细列表，在这个列表中，您可以添加、修改、删除产品和规格映射，调整程序预定义的资源映射逻辑，从而获取您真实需要的产品规格列表。

步骤1 在上云成本分析（凭证）任务列表，单击**任务名称**。

步骤2 单击操作列的“**查看配置**”按钮，进入“**产品映射**”页面。



步骤3 单击映射类型前的  按钮，可以查看映射详情。

说明

- 映射：指在不同产品或规格之间建立的对应关系。
- 标准映射：指迁移中心推荐的源端产品规格到华为云的对应产品规格。
- 非标准映射：指在没有标准映射产品规格可用时，迁移中心选择的最接近的替代产品规格，或者用户根据需求手动更改后所选择的产品规格。

----结束

产品对比

在使用成本方面，系统会对指定时间内，源端云厂商使用费用和华为云使用费用进行比较，您可以在**产品对比**页面的**对比结果**中查看。

我们还会给出源端云厂商和华为云同等规格配置的产品月均成本比较，您可以在**产品对比**页面的**产品价格对比**中查看。并且可以通过和您的销售经理沟通，针对每种类别产品分别调整价格计算比例（优惠折扣）进一步优化华为云的使用成本。

步骤1 在上云成本分析（凭证）任务列表，单击**任务名称**。

步骤2 单击操作列的“**产品对比**”按钮，进入“**产品对比**”页面。



步骤3 查看成本对比结果和**产品价格对比**，根据实际情况调整价格计算比例（优惠折扣）优化华为云的使用成本。

单击页面右上角“**导出结果**”按钮，可以将产品对比结果导出PDF格式到本地。

----结束

2.3.2 上云成本分析（账单）

通过上传源端云厂商账单导出件，可以自动完成源端云厂商消费账单分析，根据源端产品规格及用量自动匹配华为云上的产品及用量，并对比源端云厂商使用费用和华为云使用费用，为您云间迁移提供参考。

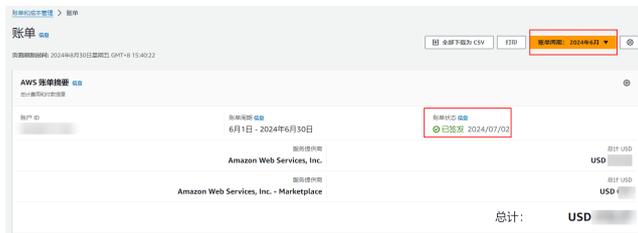
约束与限制

- 目前支持通过账单进行上云成本分析的源端云厂商有：**AWS**。
- 当前支持 EC2/RDS/S3 等17种产品的自动映射，不支持的产品将不会进行产品映射和统计。
 - Amazon EC2 (Compute)
 - Amazon EC2 (EBS)
 - Amazon EC2 (NAT Gateway)
 - Amazon RDS (MySQL)
 - Amazon RDS (AuroraMySQL)
 - Amazon EKS (Fargate)
 - Amazon ECS (Fargate)
 - Amazon DataTransfer
 - Amazon VPC (VPN)
 - Amazon VPC (VPCEP)
 - Amazon ElastiCache
 - Amazon DocDB
 - Amazon ELB
 - Amazon S3
 - Amazon DynamoDB
 - Amazon CloudWatch
 - Amazon ECR

- 每个MgC项目下，最多创建100个分析任务，且任务不能重名。
- 上传的账单文件必须满足如下要求：
 - 格式必须为英语或葡萄牙语的文字类型PDF，不支持图片类型的PDF。
 - 文件大小不超过50 MB。
 - 文件名中不能包含空格。

准备工作

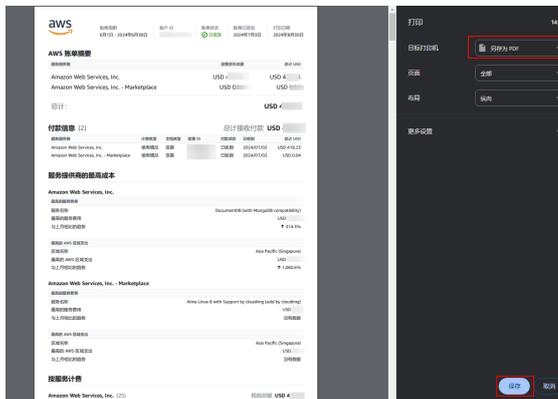
- **创建项目：**
在MgC控制台创建独立的项目，方法请参考[项目管理](#)。
- **下载AWS账单**
 - a. 打开 **AWS Billing and Cost Management 控制台**。
 - b. 在账单页面，选择需要下载账单的月份。在 AWS 账单摘要，确认账单状态显示为“已签发”。



- c. 单击页面右上角的“打印”按钮，勾选需要打印的内容，其中**最高成本、按服务计费**为必选项，单击“打印”。



- d. 在弹出的打印窗口，选择另存为PDF，单击“保存”。



创建分析任务

步骤1 登录**迁移中心**管理控制台。

步骤2 单击左侧导航栏“**调研>上云调研**”，进入“**上云调研**”页面。在页面左上角的**当前项目**下拉列表中选择在完成**准备工作**时创建的项目。

步骤3 选中**上云成本分析（账单）**框，单击“**创建分析任务**”按钮，首次使用该功能时需要授权委托，同意授权后右侧弹出“**创建分析任务**”窗口。



步骤4 自定义任务名称，并选择目的端所在区域后，单击“**添加文件**”，将导出的账单文件上传至MgC。



步骤5 账单上传成功后，单击“**确认**”按钮，提交上云成本分析任务。提交成功后可以在**上云成本分析（账单）任务列表**中查看任务状态、源端总成本和华为云总成本等详细信息。



当任务状态为“**已完成**”时，可以进行如下操作：

- **查看产品映射详情**：可以查看账单内包含的产品类型、数量以及每个产品的映射状态和结果。
- **修改映射结果**：支持手动调整产品映射结果。
- **调整价格计算比例**：通过调整华为云产品的优惠折扣，进一步优化华为云的使用成本。
- **导出成本分析结果**：将分析结果明细下载到本地进行手动调整和分析。

----**结束**

查看产品映射详情

您可以在成本分析任务完成后，查看账单内包含的所有产品类型、数量以及每个产品的映射状态和结果。

在上云成本分析（账单）任务列表，单击**任务名称**，进入任务详情页面。在任务详情页面，即可查看产品映射详情。

图 2-2 产品映射详情



修改映射结果

对于映射成功或失败的产品，均支持手动修改目的端产品类型、目的端产品和规格、目的端使用量等参数，从而调整目的端月度成本。

步骤1 在上云成本分析（账单）任务列表，单击**任务名称**，进入任务详情页面。

步骤2 在产品映射列表，单击操作列的“**修改**”，右侧弹出修改窗口。

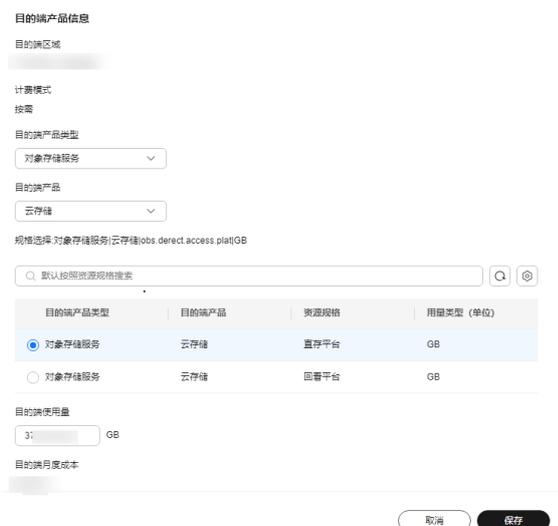
图 2-3 修改映射结果

源端产品	源端产品	数量	源端费用	源端规格	映射状态	目标端产品	目标端产品	规格	使用量	计费模式	费用 (美元)	目标端费用	目标规格	费用 (美元)	操作
Data Tran	Data Tran	50 GB	1,878,205	\$ 18.7	成功映射	Cloud Co.	Global Co.	--	--	按量	\$ 9.00	0	\$ 9.00	\$ 9.00	修改
Simple Store	Simple Store	50 GB	374,133.3	\$ 9.9	成功映射	Cloud St.	Cloud St.	64GB	274	按量	\$ 9.00	274	\$ 9.00	\$ 9.00	修改
Elastic Cl.	Elastic Cl.	50 GB	10,879.71	\$ 7.4	映射失败	--	--	--	--	按量	\$ 9.00	0	\$ 9.00	\$ 9.00	修改
Data Tran	Data Tran	50 GB	45,387.82	\$ 4.2	成功映射	--	--	--	--	按量	\$ 9.00	0	\$ 9.00	\$ 9.00	修改

步骤3 在目的端产品信息区域，根据实际需求，选择目的端产品类型、目的端产品和规格，系统会根据您的修改，重新计算目的端月度成本。

确认调整目的端产品类型和规格后，单击“**保存**”按钮。

图 2-4 修改目的端规格



----结束

调整价格计算比例

您可以通过和华为云商务代表沟通，针对每种类别产品分别调整价格计算比例（优惠折扣），进一步优化华为云的使用成本。

步骤1 在上云成本分析（账单）任务列表，单击**任务名称**，进入任务详情页面。

步骤2 在页面右上角，单击折扣后的“**修改**”，右侧弹出折扣设置窗口。

图 2-5 修改折扣比例



步骤3 根据从华为云商务代表处获得的实际折扣，调整各产品的优惠折扣。折扣设置包括两部分：总折扣和分类产品折扣。

- **总折扣：**修改总折扣会同步修改所有产品的折扣。您可以先调整总折扣，然后根据需要单独调整每个分类产品的折扣。
- **分类产品折扣设置：**根据不同类型产品的实际折扣设置，修改分类产品折扣后不会同步修改总折扣。在计算最终折扣价格时，将使用分类产品的折扣。

步骤4 所有产品折扣设置完成后，单击“**保存**”按钮。系统会根据设置的折扣计算出华为云最终的使用成本。

----结束

导出成本分析结果

您可以在成本分析任务完成后，将分析结果明细下载到本地进行手动调整和分析。

- **方法一：**在上云成本分析（账单）任务列表，单击任务操作列的“导出”，下载该任务的分析结果明细（Excel文件）。



- **方法二：**在任务详情页面，单击页面右上角的“导出”按钮，下载该任务的分析结果明细（Excel文件）。



2.3.3 问卷调研

概述

通过问卷调研的方式，基于华为云专业服务多年的经验积累，为您提供全面的源端架构梳理及资源分析功能。

运维架构调研

- 步骤1** 登录[迁移中心管理控制台](#)。
- 步骤2** 单击左侧导航栏“调研 > 上云调研”，进入“上云调研”页面。在页面左上角的当前项目下拉列表中选择[迁移项目](#)。
- 步骤3** 选中问卷调研框，单击“运维架构调研”框中的“开始”按钮，进入运维架构调研问答页面。

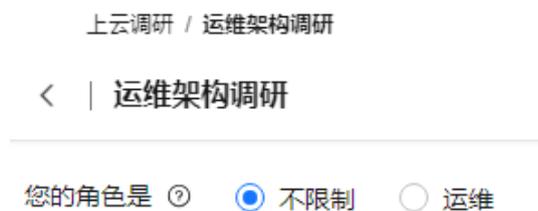
图 2-6 开始调研

问卷调研



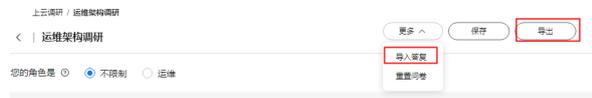
- 步骤4** 根据实际情况，选择您的角色。选择角色仅用于筛选问题，不会影响文件内容。

图 2-7 选择角色



步骤5 依次填写调研问卷所包含的问题。可以参考页面右侧的回答样例。如果在填写问卷时有疑问，或者需要咨询专业团队的意见，可以单击页面右上角的“导出”按钮，导出问卷，待专业团队进行分析并答复后，再导入问卷。

图 2-8 导出、导入问卷



步骤6 确定所有问卷问题的答复后，单击页面右上角的“保存”按钮，运维架构调研完成。

----结束

2.3.4 存储迁移成本调研

迁移中心 MgC提供了存储迁移成本评估功能，通过输入源端文件数量和大小、设定期望迁移时长等参数，系统会自动计算迁移所需的集群规格、网络带宽和迁移成本，并给出不同的迁移方案，用户可以根据项目需求和预算选择合适的迁移策略。

操作步骤

步骤1 登录[迁移中心管理控制台](#)。

步骤2 单击左侧导航栏“调研>上云调研”，进入“上云调研”页面。在页面左上角的当前项目下拉列表中选择[迁移项目](#)。

步骤3 单击[存储迁移成本调研](#)框，进入存储迁移成本调研参数配置页面。



步骤4 在源端详情区域，根据实际情况，选择源端平台和源端类型，并输入源端文件个数、总文件大小。

步骤5 在目的端详情区域，选择目的端区域和目的端类型。

步骤6 在高级特性选择网络模式、偏好设置，并输入期望迁移时间。

步骤7 配置完以上参数后，单击“创建评估”按钮。系统会根据配置参数，并给出不同的迁移方案。

----结束

2.4 资源采集

2.4.1 复杂项目

2.4.1.1 公网发现与采集

通过公网在线发现并采集源端云平台资源，获取源端资源详细信息，并整理源端资源与应用间的关联关系。目前支持发现的源端云平台为：**阿里云、华为云、AWS、腾讯云、谷歌云、Azure、七牛云、金山云**。

- **阿里云和华为云**支持采集的资源类型包括：**主机、容器、中间件、数据库、网络和存储**。
- **腾讯云**支持采集的资源类型包括：**主机、数据库、存储**。
- **谷歌云、AWS和Azure**支持采集的资源类型包括：**主机、容器、中间件、数据库、存储和网络**。
- **七牛云和金山云**支持采集的资料类型包括：**对象存储**。

创建公网发现任务

步骤1 登录**迁移中心管理控制台**。

步骤2 单击左侧导航栏的“**调研>资源采集**”，进入**资源采集**页面。在页面左上角的**当前项目**下拉列表中选择**迁移项目**。

步骤3 单击**任务框**中的“**发现>公网发现**”，进入**创建公网发现任务**页面。

图 2-9 公网发现



步骤4 根据**表2-4**，配置公网发现任务参数。

表 2-4 公网发现任务参数说明

区域	参数	说明	是否必选
基础配置	任务名称	自定义发现任务名称。	是
	任务说明	输入发现任务说明。	否
任务配置	源端	选择源端云平台，目前支持的云平台有 阿里云、华为云、AWS、腾讯云、谷歌云、Azure、七牛云、金山云 。	是

区域	参数	说明	是否必选
	凭证	<p>选择源端对应的凭证，如未提前添加凭证，请单击“新建凭证”，填写相应参数，校验并保存凭证。</p> <ul style="list-style-type: none"> • 阿里云、华为云、AWS、腾讯云、七牛云、金山云认证方式请选择“AK/SK”，并输入源端账号的AK/SK。 • 谷歌云认证方式请选择“配置文件”，并上传谷歌云服务账号凭据，该凭证必须为json格式文件，且大小不能超过4KB。 • Azure认证方式请选择“ID/密钥”。添加凭证所需的信息获取方法请参见如何获取添加Azure凭证所需的信息。 	是
	区域	选择要发现的源端云平台区域，支持多选。	是

步骤5 开启云平台采集，在**资源类型**下拉列表，勾选需要进行采集的资源类型（采集项）。当前支持采集的云平台和资源类型参见[表2-5](#)。

图 2-10 选择资源类型



表 2-5 支持采集的资源类型

云平台	资源类型	子类型
阿里云	<ul style="list-style-type: none"> • 主机 • 容器 • 大数据 • 数据库 	-
	中间件	<ul style="list-style-type: none"> • Redis • Kafka
	存储	<ul style="list-style-type: none"> • 对象存储 • 文件存储

云平台	资源类型	子类型
	网络	<ul style="list-style-type: none"> 云连接 负载均衡（ALB、CLB） 专线 公网域名 内网域名 弹性公网IP 公网NAT网关 路由表 安全组 虚拟私有云 VPN网关
华为云	<ul style="list-style-type: none"> 主机 容器 大数据 数据库 	-
	中间件	<ul style="list-style-type: none"> Redis Kafka
	存储	<ul style="list-style-type: none"> 对象存储 文件存储
	网络	<ul style="list-style-type: none"> 负载均衡（ELB） 公网域名 内网域名 弹性公网IP 公网NAT网关 路由表 安全组 虚拟私有云
AWS	<ul style="list-style-type: none"> 主机 容器 数据库 	-
	中间件	<ul style="list-style-type: none"> Redis Kafka

云平台	资源类型	子类型
	网络	<ul style="list-style-type: none"> ● 负载均衡（ELB）（目前只支持CLB类型） ● 公网域名 ● 内网域名 ● 弹性公网IP ● 公网NAT网关 ● 路由表 ● 安全组 ● 虚拟私有云
	存储	<ul style="list-style-type: none"> ● 对象存储 ● 文件存储
腾讯云	<ul style="list-style-type: none"> ● 主机 ● 数据库 	-
	存储	<ul style="list-style-type: none"> ● 对象存储 ● 文件存储
谷歌云	<ul style="list-style-type: none"> ● 主机 ● 容器 ● 数据库 	-
	中间件	Redis
	存储	<ul style="list-style-type: none"> ● 对象存储 ● 文件存储
	网络	<ul style="list-style-type: none"> ● 弹性公网IP ● 路由表 ● 安全组 ● 虚拟私有云
Azure	<ul style="list-style-type: none"> ● 主机 ● 容器 ● 数据库 	-
	存储	<ul style="list-style-type: none"> ● 对象存储 ● 文件存储
	中间件	<ul style="list-style-type: none"> ● Redis ● Kafka

云平台	资源类型	子类型
	网络	<ul style="list-style-type: none"> 弹性公网IP 路由表 安全组 公网NAT网关 虚拟私有云 负载均衡（ELB）
七牛云	存储	对象存储
金山云	存储	对象存储

步骤6 （可选）选择需要采集的源端云平台服务（仅支持阿里云、腾讯云），获取资源与服务/微服务间的调用关系、关联关系以及依赖关系。

须知

选择采集的服务时，需要确保所提供的凭证拥有该服务的相关权限。

采集各服务的功能见下表。

服务	功能
资源管理 RM	通过采集该服务可以发现更多应用与资源间的关联线索，防止应用与资源间关系链断裂。
域名解析服务 DNS	通过采集该服务可以从接入层分析应用间的关联关系。
应用防火墙 WAF	通过采集该服务可以获得接入层的流量方向。
负载均衡 SLB	通过采集该服务可以获得接入层的流量方向。
微服务引擎 MSE	通过采集该服务可以获得应用间的关联关系，微服务间的调用关系。

步骤7 单击“确认”，公网发现任务创建完成，系统自动开始资源发现。

- 您可以在[资源采集](#)页面，[查看发现的源端资源列表和资源详情](#)。
- 您可以在[任务列表](#)页面，[查看任务状态和详情](#)。

步骤8 发现任务状态为成功后，根据不同资源类型，进行后续操作。

- 主机**资源需要先完成[深度采集](#)或[源端迁移准备度检查](#)，再进行[迁移方案设计](#)或[迁移方案配置](#)。
- 容器**资源需要先完成[深度采集](#)，再进行[迁移方案设计](#)。
- 中间件、数据库、存储**等资源可直接进行[迁移方案设计](#)。支持对AWS RDS（包括MySQL、MariaDB、Aurora、PGSQL、SQLServer、Oracle）和AWS DocumentDB数据库进行[深度采集](#)，以获取数据库更详细的信息。

- 对象存储资源可以进行[深度采集](#)和[存储迁移方案配置](#)。

----结束

主机源端迁移准备度检查

请按照以下步骤对主机资源进行迁移预检查（源端迁移准备度），确定源端主机是否满足迁移条件。

1. 在源端内网环境中[安装Edge工具](#)并完成注册。
2. 在[资源采集](#)页面的[资源列表](#)，单击主机或主机的资源数量，进入[资源列表](#)页面的[主机](#)页签。

资源列表

资源类型	资源数量
主机	138
容器	0
中间件	2
数据库	1
大数据	0
网络	496
存储	116

3. 勾选待迁移的主机，单击列表上方的[关联到应用](#)按钮，弹出关联到应用窗口。
 - 已提前[创建应用](#)，在下拉列表中选择要关联的应用，单击“[确定](#)”。
 - 未提前创建应用，单击“[新建应用](#)”，输入自定义的[应用名称](#)和[描述](#)，选择[业务场景](#)、[使用场景](#)和[目的端区域](#)，单击“[创建](#)”按钮，应用创建成功。选择创建的应用后，单击“[确定](#)”。
4. 在列表上方选择“[目标场景](#)>[主机迁移](#)”，此时源端迁移准备度列会出现“[配置](#)”按钮。
5. 单击源端迁移准备度列的“[配置](#)”，弹出配置窗口。
6. 根据[表2-6](#)，配置参数。

表 2-6 迁移准备度参数配置说明

参数	配置说明
类型	根据实际的主机操作系统类型选择。
采集设备	选择在源端环境安装的Edge工具。
接入IP	选择主机接入IP，可以为公网IP，也可以是私有IP；迁移预检查通过后，该IP会作为后续主机迁移工作流的资源接入IP。
端口	主机开放端口。 <ul style="list-style-type: none"> ● Windows主机默认为5985端口，无法修改。 ● Linux主机默认22端口，可根据实际情况修改。

参数	配置说明
凭证	<p>选择主机凭证，如未提前在Edge上添加源端主机凭证，请前往Edge控制台添加，并同步至迁移中心。</p> <p>须知 为确保采集过程具备充分的系统访问权限，从而能够获取到必要的信息和数据。对主机深度采集的凭证要求如下：</p> <ul style="list-style-type: none"> 对Linux主机进行深度采集时，请添加Root账号和密码作为采集凭证。 对Windows主机进行深度采集时，请添加Administrator账号和密码作为采集凭证。

- 配置完成后，单击“**确认**”按钮，系统会校验配置信息的正确性并开始源端迁移准备度检查，当源端迁移准备度列状态会变为**已就绪**时，代表该主机已满足迁移条件，可以进行下一阶段[迁移方案设计](#)或[迁移方案配置](#)。

主机深度采集

请按照以下步骤对**主机资源**进行深度采集。

- 在源端内网环境中[安装Edge工具](#)并与迁移中心连接。
- 在**资源采集**页面的**资源列表**，单击主机资源的**资源数量**，进入**资源列表**页面的**主机**页签。
- 在需要进行深度采集的主机**采集设备**列，单击“**配置**”，弹出**配置**窗口。根据表 2-7，配置参数。

表 2-7 深度采集参数配置说明

参数	配置说明
类型	根据实际的主机操作系统类型选择。
采集设备	选择在源端环境安装的Edge工具。
接入IP	选择主机接入IP，可以为公网IP，也可以是私有IP；迁移预检查通过后，该IP会作为后续主机迁移工作流的资源接入IP。
端口	<p>主机开放端口。</p> <ul style="list-style-type: none"> Windows主机默认为5985端口，无法修改。 Linux主机默认22端口，可根据实际情况修改。
凭证	<p>选择主机凭证，如未提前在Edge上添加源端主机凭证，请前往Edge控制台添加，并同步至迁移中心。</p> <p>须知 为确保采集过程具备充分的系统访问权限，从而能够获取到必要的信息和数据。对主机深度采集的凭证要求如下：</p> <ul style="list-style-type: none"> 对Linux主机进行深度采集时，请添加Root账号和密码作为采集凭证。 对Windows主机进行深度采集时，请添加Administrator账号和密码作为采集凭证。

4. 单击“**确认**”，采集设备和凭证都绑定成功后，系统会自动开始深度采集。当深度采集列的状态为“**已采集**”时，代表采集完成。可以进行下一阶段[迁移方案设计](#)或[迁移方案配置](#)。

容器深度采集

请按照以下步骤对**容器资源**进行深度采集。

1. 在源端内网环境中[安装Edge工具](#)并与迁移中心连接。
2. 在**资源采集**页面的**资源列表**，单击容器资源的**资源数量**，进入**资源列表**页面的**容器**页签。
3. 在需要进行深度采集的**容器采集设备**列，单击“**绑定**”，弹出**绑定采集设备**窗口。
如果需要批量绑定，可以同时勾选多个资源，单击列表上方的“**绑定采集设备**”，弹出**绑定采集设备**窗口。
4. 选择源端安装的Edge工具，单击“**确定**”。绑定成功后，资源的采集设备列状态会变为已绑定。
5. 采集设备绑定成功后，还需要为资源绑定凭证，单击资源凭证列的“**绑定**”，弹出**绑定凭证**窗口。
6. 选择资源凭证，如未提前在Edge上[添加源端资源凭证](#)，请前往Edge控制台添加，并同步至迁移中心。
7. 单击“**确定**”，系统会自动检查凭证绑定状态。当深度采集列状态为**就绪**时，单击深度采集列的“**采集**”进行深度采集。首次深度采集完成后，可以单击采集状态列的“**重新采集**”按钮，进行多次深度采集。采集完成后，单击资源名称可以查看采集到的容器详情。

对象存储深度采集

通过深度采集获取对象存储资源的详细信息，以提高迁移集群规格评估的准确性。请按照以下步骤进行**对象存储资源**深度采集。

1. 在源端内网环境中[安装Edge工具](#)并与迁移中心连接。
2. 在**资源采集**页面的**资源列表**，单击**存储资源**的**资源数量**，进入**资源列表**页面的**存储**页签。
3. 在对象存储资源列表中，单击**深度采集配置**列的“**配置**”，弹出**配置**窗口。
4. 选择源端安装的Edge工具和相应的资源凭证，单击“**确认**”完成配置。如未提前在Edge上[添加源端资源凭证](#)，请前往Edge控制台添加，并同步至迁移中心。

须知

在对**Azure**云平台对象存储资源进行深度采集时，需要使用“**存储账户**”和“**密钥**”作为采集凭证，凭证的获取方法请参见[如何获取Azure对象存储深度采集所需凭证](#)。

5. 配置完成后，单击操作列的“**添加前缀**”，弹出**添加前缀**窗口。
6. 输入指定的资源目录路径进行采集，不输入则默认采集全桶资源，单击“**确定**”保存前缀设置。
7. 完成深度采集配置和添加前缀后，单击操作列的“**深度采集**”，系统开始进行对象存储资源的深度采集。支持进行多次深度采集。当采集状态为**采集完成**后，单击资源名称，可以查看采集到的信息。

数据库深度采集

当前支持对AWS RDS（包括MySQL、MariaDB、Aurora、PostgreSQL、SQL Server、Oracle）和AWS DocumentDB数据库进行深度采集，以获取包括数据库版本、引擎、服务器字符集、平均事务每秒(Transaction Per Second, TPS)和查询每秒(Query Per Second, QPS)等关键性能指标在内的详细信息。不同的数据库类型，采集的信息详情可能有所差异。

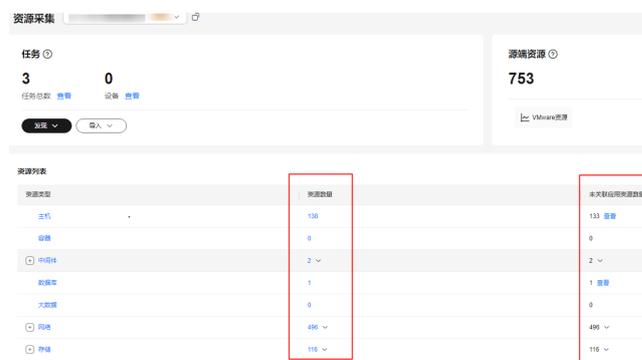
1. 在源端内网环境中或在可以访问到源端数据库的网络环境中[安装Edge工具](#)并与迁移中心连接。
2. 在[资源采集](#)页面的[资源列表](#)，单击数据库资源的[资源数量](#)，进入[资源列表](#)页面的[数据库](#)页签。
3. 在数据库资源列表中，通过[平台类型](#)过滤出采集到的所有AWS数据库资源，在支持深度采集的数据库类型的[采集设备](#)列，单击“[绑定](#)”，弹出[绑定采集设备](#)窗口。
如果需要批量绑定，可以同时勾选多个资源，单击页面右上角的“[绑定采集设备](#)”，弹出[绑定采集设备](#)窗口。
4. 选择源端安装的Edge工具，以及网络接入方式（如果采集的资源与安装Edge的主机处于同一VPC内，可以选择私网接入，否则需要选择公网接入），单击“[确定](#)”。绑定成功后，资源的采集设备列状态会变为已绑定。
5. 采集设备绑定成功后，还需要为资源绑定凭证，单击资源[凭证](#)列的“[绑定](#)”，弹出[绑定凭证](#)窗口。
6. 选择数据库资源凭证，如未提前在Edge上[添加源端资源凭证](#)，请前往Edge控制台添加，并同步至迁移中心。
7. 单击“[确定](#)”，系统会自动检查凭证绑定状态。当[深度采集](#)列的状态为[就绪](#)时，单击“[采集](#)”开始深度采集。首次深度采集完成/采集失败后，可以单击深度采集列的“[重新采集](#)”，进行多次深度采集。
8. 深度采集完成后，单击数据库名称，进入[数据库详情](#)页面，在[数据库详情](#)区域，可以查看采集到的详细信息。

查看发现的源端资源列表和资源详情

发现任务仅能获取源端资源的基本信息，更多详细信息需要通过深度采集获取。

1. 在[资源采集](#)页面的“[资源列表](#)”，可以查看发现的各类型资源数量和未关联应用的资源数量。

图 2-11 查看资源列表



2. 单击资源类型或资源数量，进入对应的资源列表。

3. 单击资源名称，可以查看该资源的基本信息。

查看任务状态和详情

步骤1 在资源采集页面，单击任务总数后的“查看”，进入任务列表页面。

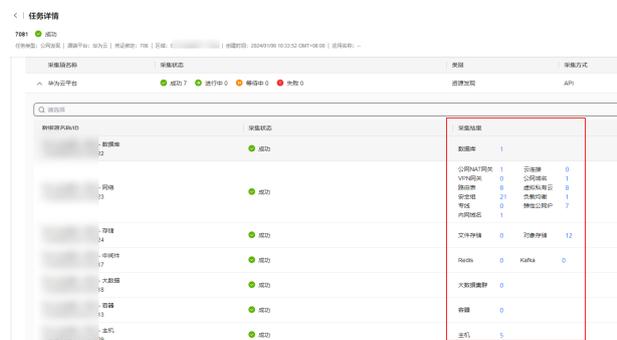
图 2-12 查看任务列表



步骤2 单击采集任务操作列的“查看”，进入任务详情页面，可以查看到该任务包含的采集项。

步骤3 单击采集项前的 ✓ 按钮，可以查看该采集项包含的采集数据源以及各资源类型的采集结果。

图 2-13 采集结果



如果有采集失败的数据源，可以将鼠标放到数据源的采集状态上，查看失败原因。



步骤4 在采集结果列，单击资源类型后的数量，可以跳转到对应的资源类型列表页面。

----结束

2.4.1.2 内网发现与采集

在源端内网环境中部署Edge工具，通过**网段扫描**或**VMware扫描**的方式，发现和采集源端主机资源。

使用须知

- 目前支持采集的VMware版本包括：**VMware 5.0~7.0**
- VMware扫描和网段扫描两种方法都采用私有IP结合Edge ID的方式来标识内网中的主机资源。这意味着，一旦完成一次资源采集，如果被采集的主机私有IP发生变更，系统将在下一次采集时将其识别为一个新的主机资源，从而导致主机资源数量增加。为了避免这种情况，建议在迁移过程完成之前，不要更改源端主机的私有IP地址。

前提条件

- 已在源端内网环境**安装Edge工具**，并与迁移中心连接。
- 已在Edge上**添加源端主机凭证**。

须知

为确保采集过程具备充分的系统访问权限，从而能够获取到必要的信息和数据。对主机深度采集的凭证要求如下：

- 对Linux主机进行深度采集时，请添加Root账号和密码作为采集凭证。
- 对Windows主机进行深度采集时，请添加Administrator账号和密码作为采集凭证。

创建内网采集任务

步骤1 登录**迁移中心管理控制台**。

步骤2 单击左侧导航栏的“**调研>资源采集**”，进入**资源采集**页面。在页面左上角的**当前项目**下拉列表中选择**迁移项目**。

步骤3 单击任务框中的“**发现>内网发现**”，进入**创建内网发现任务**页面。

图 2-14 内网发现



步骤4 根据表2-8，配置内网发现任务参数。

表 2-8 内网发现任务参数说明

参数	说明
任务名称	自定义发现任务名称。
任务说明	输入发现任务说明。
设备	选择在源端内网环境中安装的Edge工具。

步骤5 根据实际情况，选择使用网段扫描或VMware扫描发现主机资源。

- 选择网段扫描，根据表2-9，配置网段扫描参数。

表 2-9 网段扫描参数说明

参数	说明
协议	选择通信协议TCP或ICMP。
网段	输入IP地址段，需在以下范围内： - 10.0.0.0 ~10.255.255.255 - 172.16.0.0 ~172.31.255.255 - 192.168.0.0 ~ 192.168.255.255
Linux	填写Linux系统类型主机需要扫描的端口。仅TCP协议需要填写，若不需要扫描Linux类型主机，可将端口设为0。
Windows	填写Windows系统类型主机需要扫描的端口。仅TCP协议需要填写，若不需要扫描Windows类型主机，可将端口设为0。

- 选择VMware扫描，在连接地址填写vCenter的IP地址，并选择vCenter的凭证。如未提前添加凭证，请前往迁移工具Edge添加私有云凭证。凭证中的用户名、密码为vCenter的登录用户名和密码。

步骤6 单击“确认”，内网发现任务创建完成，系统开始主机资源发现。

在资源采集页面，单击任务总数后的“查看”，进入任务列表页面，查看任务状态。

步骤7 内网发现任务状态为成功后，即可进行主机深度采集。

- 在资源采集页面的资源列表，单击主机资源的资源数量，进入资源列表页面的主机页签。
- 单击主机采集设备列表或凭证列的“配置”，弹出配置窗口。
- 根据表2-10，配置参数。

表 2-10 深度采集参数配置说明

参数	配置说明
类型	根据实际的主机操作系统类型选择。
采集设备	选择在源端环境安装的Edge工具。
接入IP	选择主机接入IP，可以为公网IP，也可以是私有IP；迁移预检查通过后，该IP会作为后续主机迁移工作流的资源接入IP。
端口	主机开放端口。 - Windows主机默认为5985端口，无法修改。 - Linux主机默认22端口，可根据实际情况修改。
凭证	选择主机凭证，如未提前在Edge上 添加源端主机凭证 ，请前往Edge控制台添加，并同步至迁移中心。

- 单击“确认”，采集设备和凭证都绑定成功后，系统会自动开始深度采集。当深度采集列的状态为“已采集”时，代表采集完成。可以进行下一阶段[迁移方案设计](#)或[迁移方案配置](#)。

图 2-15 采集完成



----结束

2.4.1.3 导入应用关联文件

将您源端配置管理数据库中的应用关联信息，填入导入模板表格中，再将表格导入迁移中心，分析出应用间的依赖及关联关系。

操作步骤

- 步骤1 登录[迁移中心管理控制台](#)。
- 步骤2 单击左侧导航栏的“调研>资源采集”，进入资源采集页面。在页面左上角的当前项目下拉列表中选择[迁移项目](#)。
- 步骤3 单击任务框中的“导入>应用关联模板”，弹出导入应用关联模板窗口，单击“下载模板”，将导入模板下载到本地。

图 2-16 下载应用关联模板



步骤4 打开导入模板，填写模板内的参数，其中标黄的参数为必填。模板参数填写完成并保存后，单击导入窗口的“添加文件”，将填写后的模板文件上传至迁移中心。

步骤5 上传成功后，单击“确认”按钮，系统自动开始采集。

您可以在任务列表页面，查看任务状态。

- 如果任务状态为“失败”，单击操作列的“查看”，可以查看采集失败的数据源，将鼠标放到数据源的采集状态上，可以查看失败原因。处理完失败原因后，需要删除该采集任务，重新导入文件进行采集。
- 当任务状态为“成功”时，您可以前往应用管理页面，[查看应用关联分析结果](#)。

----结束

2.4.1.4 导入工具采集结果

通过线下采集器，采集本地资源信息，并将采集结果上传到迁移中心进行应用关联分析。

前提条件

已[创建工具采集任务](#)，并获取采集结果。

操作步骤

步骤1 登录[迁移中心管理控制台](#)。

步骤2 单击左侧导航栏的“调研>资源采集”，进入资源采集页面。在页面左上角的当前项目下拉列表中选择[迁移项目](#)。

步骤3 单击任务框中的“导入>工具采集结果”，弹出导入工具采集结果窗口。

图 2-17 导入工具采集结果



步骤4 单击导入窗口的“添加文件”，将工具采集结果的json文件上传至迁移中心。

步骤5 上传成功后，单击“确认”按钮，系统自动开始采集。

您可以在任务列表页面，查看任务状态。

- 如果任务状态为“失败”，单击操作列的“查看”，可以查看采集失败的数据源，将鼠标放到数据源的采集状态上，可以查看失败原因。处理完失败原因后，需要删除该采集任务，重新导入文件进行采集。
- 当任务状态为“成功”时，您可以前往应用管理页面，[查看应用关联分析结果](#)。

----结束

2.4.1.5 导入阿里云资源清单

从阿里云导出资源清单（CSV格式），再将其导入MgC，系统会自动解析这些文件，识别出相应的资源，并在资源列表展示资源的详细信息。此外，利用解析后的信息，您还可以使用迁移中心的规格评估功能，为您推荐与阿里云资源相匹配的华为云资源规格。

支持导入的阿里云资源类型包括：

- 云服务器 ECS实例和云盘
- 对象存储 OSS桶
- 云数据库 RDS MySQL实例
- 云数据库 Redis实例

导出阿里云 ECS 实例和云盘

步骤1 登录阿里云ECS控制台。

步骤2 在概览页面，单击“导出数据>实例”，导出所有实例资源清单。



步骤3 单击“导出数据>云盘”，导出所有云盘资源清单。

步骤4 打开实例资源清单和云盘资源清单，根据实际需求，调整清单内的资源信息。请确保清单中包含了以下必要信息：

清单类型	必选字段
实例清单	实例ID、操作系统、实例规格、CPU、内存
云盘清单	实例Id、云盘Id、容量（GiB）、云盘属性

----结束

导出阿里云 OSS 桶

步骤1 登录阿里云OSS控制台。

步骤2 在左侧导航栏选择“**Bucket 列表**”，进入**Bucket 列表**页面。

步骤3 单击列表右上角的  图标，弹出**导出Bucket列表**窗口。



步骤4 勾选需要导出的信息，单击“**确定**”，导出所有Bucket资源清单。

----结束

导出阿里云 RDS MySQL 实例

步骤1 登录阿里云RDS控制台。

步骤2 在左侧导航栏选择“**实例列表**”，进入**实例列表**页面。在页面上方选择地域。

步骤3 通过**数据库类型**列过滤出所有MySQL类型的实例，单击列表右上角的  图标。



步骤4 在弹出的对话框中，勾选需要导出的信息。单击“**确定**”。系统会将实例信息输出至CSV格式的文件并自动下载。

步骤5 由于从阿里云导出的CSV文件内容格式不满足MgC解析标准，因此需要进行如下修改：

1. 使用文件编辑器打开下载的资源清单CSV文件，在表头行搜索“PrivatePort(A network port,)”。
2. 将该表头名中的英文逗号删除，然后保存文件。正确的表头名应该为“PrivatePort(A network port)”。

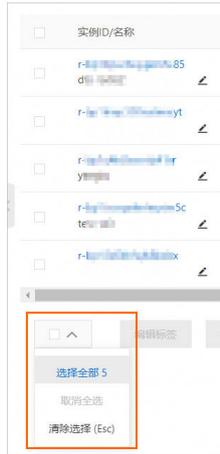
----结束

导出阿里云 Redis 实例

步骤1 登录阿里云云数据库 Redis 版控制台。

步骤2 在左侧导航栏选择“**实例列表**”，进入**实例列表**页面。在页面上方选择地域。

步骤3 在实例列表页面左下方，展开  ，单击“**选择全部**”。



步骤4 单击列表右上角的  图标，在弹出的对话框中，勾选需要导出的信息。



步骤5 单击“确定”。系统会将实例信息输出至CSV格式的文件并自动下载。

步骤6 由于从阿里云导出的CSV文件内容格式不满足MgC解析标准，因此需要进行如下修改：

1. 使用文件编辑器打开下载的资源清单CSV文件，
2. 在表头行的末尾，添加一个英文逗号，然后保存文件。

例如，原始表头为：

```
ID,Name,Quantity
```

修改后的表头应该是：

```
ID,Name,Quantity,
```

----结束

导入阿里云资源清单

步骤1 登录[迁移中心管理控制台](#)。

步骤2 单击左侧导航栏的“[调研>资源采集](#)”，进入[资源采集](#)页面。在页面左上角的当前项目下拉列表中选择[迁移项目](#)。

步骤3 单击任务框中的“[导入>阿里云资源清单](#)”，弹出[导入阿里云资源清单](#)窗口。

图 2-18 导入阿里云资源清单



须知

仅支持上传UTF-8编码且为中文表头的CSV文件，单个文件大小不能超过15MB。将CSV文件编码转换为UTF-8的方法请参考[转换CSV文件编码为UTF-8](#)。

步骤4 在资源类型下拉列表中，选择要导入的资源类型，并上传对应的资源清单。

导入阿里云资源清单

默认采用覆盖策略：若检测到相同实例ID的资源，新资源信息将自动覆盖旧资源信息。同时，新导入的关联和依赖关系将与原有关系整合更新。

* 资源类型 主机 [如何获取资源清单](#)

* 资源清单 添加文件
仅支持上传UTF-8编码且为中文表头的CSV文件，且单个文件大小不能超过50MB。
[UTF-8编码转换方法](#)

* 云盘资源清单 添加文件
仅支持上传UTF-8编码且为中文表头的CSV文件，且单个文件大小不能超过50MB。

取消 确认

资源类型	操作步骤
主机	1. 单击资源清单后的“添加文件”，将导出的阿里云ECS实例资源清单上传。 2. 实例资源清单上传后，单击云盘资源清单后的“添加文件”，将导出的阿里云ECS云盘资源清单上传。
Redis	单击资源清单后的“添加文件”，将导出的阿里云Redis实例资源清单上传。
对象存储	单击资源清单后的“添加文件”，将导出的阿里云OSS桶资源清单上传。
数据库	单击资源清单后的“添加文件”，将导出的阿里云RDS MySQL实例资源清单上传。

步骤5 单击“确认”，资源清单上传成功后，在资源列表中选择导入的资源类型，进入对应的资源列表页面，可以查看导入的资源列表。

步骤6 （可选）将导入的阿里云资源[关联应用](#)，并使用[规格评估](#)功能，获取与阿里云资源相匹配的华为云资源规格。

----结束

2.4.1.6 导入 RVTools 资源

约束与限制

- **版本要求**
支持导入RVTools以下版本的导出结果：
 - 4.4.1
 - 4.4.2
 - 4.4.3
 - 4.4.4
 - 4.4.5
 - 4.5.0
 - 4.5.1
 - 4.6.1
- **文件格式要求**
导出结果的文件格式必须为Excel (.xlsx)。
- **文件大小、压缩比要求**
导入的文件大小不能超过100 MB，压缩比不能低于5%。

步骤一：导出 RVTools 资源

步骤1 启动RVTools应用程序，进入登录界面。

步骤2 在登录界面输入以下信息。

- 在IP地址栏中，输入vCenter服务器的IP地址。
- 在用户名栏中，填写用于连接vCenter的用户名。
- 在密码栏中，输入与用户名相对应的密码。

步骤3 单击“**登录**”按钮。登录成功后，在菜单栏单击“**file > Export all to Excel**”。

步骤4 选择保存路径，将资源导出到本地。

----结束

步骤二：将导出的文件上传至 OBS

步骤1 登录华为云**OBS控制台**，创建一个用于存储RVTools导出资源文件的**标准存储桶**。方法请参考**创建桶**。

步骤2 参考**上传对象**，将**导出的RVTools资源**文件上传至OBS标准存储桶。

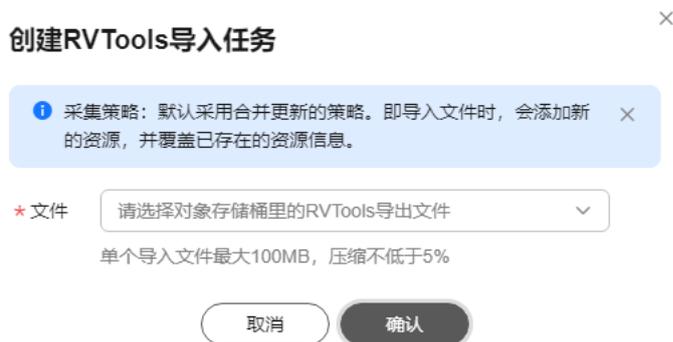
----结束

步骤三：导入 RVTools 资源

步骤1 登录**迁移中心管理控制台**。

步骤2 单击左侧导航栏的“**调研>资源采集**”，进入**资源采集**页面。在页面左上角的**当前项目**下拉列表中选择**迁移项目**。

步骤3 单击任务框中的“导入 > RVTools”，弹出创建RVTools导入任务窗口。



说明

首次使用该功能时需要授权委托，单击RVTools导入委托窗口的“同意授权”即可。

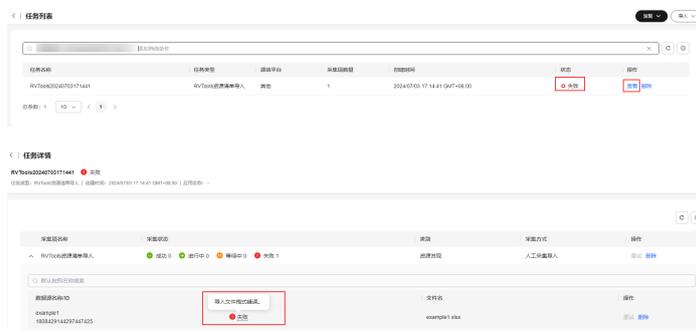
步骤4 在文件的下拉列表中，找到**步骤二**存放RVTools导出资源文件的OBS桶，并选择导出的资源文件，单击“确认”按钮，上传至迁移中心。

步骤5 上传成功后，单击任务框中，任务总数后的“查看”，进入任务列表页面，查看任务状态。



- 如果任务状态为“失败”，单击操作列的“查看”，可以查看采集失败的数据源，将鼠标放到数据源的采集状态上，可以查看失败原因。常见失败原因的处理方法请参考[导入RVTools资源任务失败](#)。

处理完失败原因后，需要删除该采集任务，重新导入文件进行采集。



- 当任务状态为“成功”时，您可以查看导入的[资源详情](#)和[资源统计信息](#)。

----结束

查看主机资源详情

当RVTools导入任务状态为“成功”时，在资源列表页面，可以查看VMware采集的主机信息，包括：主机名称、实例ID、IP地址、主机规格、操作系统类型以及平台类型等。

步骤1 在资源采集页面的资源列表，单击主机资源类型或资源数量，进入资源列表页面的主机页签。

资源列表

资源类型	资源数量
主机	138
容器	0
中间件	2
数据库	1
大数据	0
网络	496
存储	116

步骤2 在主机资源列表，可以查看导入的主机资源信息，单击资源名称，可以查看该主机的更多基本信息、规格信息等。



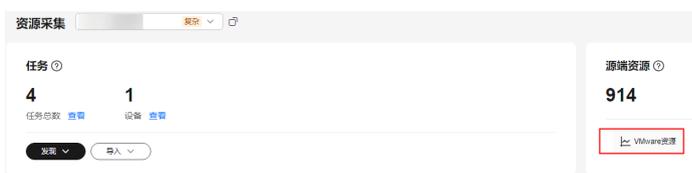
----结束

查看资源统计信息

当RVTools导入任务状态为“成功”时，在资源统计页面，可以查看VMware资源的统计信息，包括：

- 宿主机：CPU类型统计、宿主机个数统计、内存总量统计、已分配虚拟机内存、已分配占比、虚拟机总数统计、虚拟机运行总数统计、存储使用详情统计以及主机列表详情。
- 虚拟机：Windows操作系统个数统计、Linux系统个数统计、内存使用总容量、CPU使用核数、磁盘使用个数、磁盘容量占比、虚拟机数量、虚拟机列表、操作系统。

步骤1 在资源采集页面，单击源端资源框中的“VMware资源”，进入资源统计页面。



步骤2 分别选择VMware资源的“**宿主机**”页签和“**虚拟机**”页签，查看对应的统计信息。

---结束

2.4.2 简单项目

2.4.2.1 公网发现与采集

通过公网在线发现并采集源端云平台资源，获取源端资源详细信息。目前支持的源端云平台为：**阿里云、华为云、AWS、腾讯云、谷歌云、Azure、七牛云、金山云**。

- **阿里云和华为云**支持采集的资源类型包括：**主机、容器、中间件、数据库、网络和存储**。
- **腾讯云**支持采集的资源类型包括：**主机、数据库、存储**。
- **谷歌云、AWS和Azure**支持采集的资源类型包括：**主机、容器、中间件、数据库、存储和网络**。
- **七牛云和金山云**支持采集的资料类型包括：**对象存储**。

创建公网发现任务

步骤1 登录**迁移中心管理控制台**。

步骤2 单击左侧导航栏的“**调研>资源采集**”，进入**资源采集**页面。在页面左上角的当前项目下拉列表中选择**迁移项目**。

步骤3 单击**任务框**中的“**发现>公网发现**”，进入**创建公网发现任务**页面。

图 2-19 公网发现



步骤4 根据**表2-11**，配置公网发现任务参数。

表 2-11 公网发现任务参数说明

区域	参数	说明	是否必选
基础配置	任务名称	自定义发现任务名称。	是
	任务说明	输入发现任务说明。	否

区域	参数	说明	是否必选
任务配置	源端	选择源端云平台，目前支持的云平台有 阿里云、华为云、AWS、腾讯云、谷歌云、Azure、七牛云、金山云 。	是
	凭证	选择源端对应的凭证，如未提前 添加凭证 ，请单击“ 新建凭证 ”，填写相应参数，校验并保存凭证。 <ul style="list-style-type: none"> • 阿里云、华为云、AWS、腾讯云、七牛云、金山云认证方式请选择“AK/SK”，并输入源端账号的AK/SK。 • 谷歌云认证方式请选择“配置文件”，并上传谷歌云服务账号凭据，该凭证必须为json格式文件，且大小不能超过4KB。 • Azure认证方式请选择“ID/密钥”。添加凭证所需的信息获取方法请参见如何获取添加Azure凭证所需的信息。 	是
	区域	选择要发现的源端云平台区域，支持多选。	是

步骤5 在**资源类型**下拉列表，勾选需要进行采集的资源类型（采集项）。当前支持采集的云平台和资源类型参见**表2-12**。

图 2-20 选择要采集的资源类型

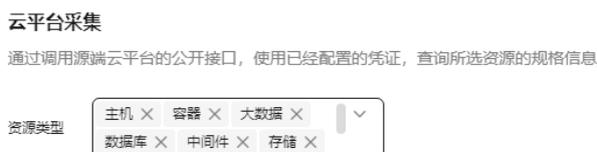


表 2-12 支持采集的资源类型

云平台	资源类型	子类型
阿里云	<ul style="list-style-type: none"> • 主机 • 容器 • 大数据 • 数据库 	-

云平台	资源类型	子类型
	中间件	<ul style="list-style-type: none"> Redis Kafka
	存储	<ul style="list-style-type: none"> 对象存储 文件存储
	网络	<ul style="list-style-type: none"> 云连接 负载均衡（ALB、CLB） 专线 公网域名 内网域名 弹性公网IP 公网NAT网关 路由表 安全组 虚拟私有云 VPN网关
华为云	<ul style="list-style-type: none"> 主机 容器 大数据 数据库 	-
	中间件	<ul style="list-style-type: none"> Redis Kafka
	存储	<ul style="list-style-type: none"> 对象存储 文件存储
	网络	<ul style="list-style-type: none"> 负载均衡（ELB） 公网域名 内网域名 弹性公网IP 公网NAT网关 路由表 安全组 虚拟私有云
AWS	<ul style="list-style-type: none"> 主机 容器 数据库 	-

云平台	资源类型	子类型
	存储	<ul style="list-style-type: none"> 对象存储 文件存储
	网络	<ul style="list-style-type: none"> 负载均衡（ELB）（目前只支持CLB类型） 公网域名 内网域名 弹性公网IP 公网NAT网关 路由表 安全组 虚拟私有云
	存储	<ul style="list-style-type: none"> 对象存储 文件存储
腾讯云	<ul style="list-style-type: none"> 主机 数据库 	-
	存储	<ul style="list-style-type: none"> 对象存储 文件存储
谷歌云	<ul style="list-style-type: none"> 主机 容器 数据库 	-
	中间件	Redis
	存储	<ul style="list-style-type: none"> 对象存储 文件存储
	网络	<ul style="list-style-type: none"> 弹性公网IP 路由表 安全组 虚拟私有云
Azure	<ul style="list-style-type: none"> 主机 容器 数据库 	-
	存储	<ul style="list-style-type: none"> 对象存储 文件存储
	中间件	<ul style="list-style-type: none"> Redis Kafka

云平台	资源类型	子类型
	网络	<ul style="list-style-type: none"> 弹性公网IP 路由表 安全组 公网NAT网关 虚拟私有云 负载均衡（ELB）
七牛云	存储	对象存储
金山云	存储	对象存储

步骤6（可选）将采集的资源关联到应用。

- 已提前[创建应用](#)，在“应用”的下拉列表中选择要关联的应用。
- 未提前创建应用，单击“新建应用”，弹出新建应用窗口，输入自定义的**应用名称**和**描述**；根据实际需求，选择**业务场景和使用场景**；**区域**选择目的端所在区域，单击“确定”按钮，应用创建成功。

步骤7 单击“确认”，公网发现任务创建完成，系统自动开始资源发现。

- 您可以在[资源采集](#)页面，[查看发现的源端资源列表和资源详情](#)。
- 您可以在[任务列表](#)页面，[查看任务状态和详情](#)。

步骤8 发现任务状态为成功后，根据不同资源类型，进行后续操作。

- **主机资源**需要先完成[深度采集](#)或[源端迁移准备度检查](#)，再进行[迁移方案设计](#)或[迁移方案配置](#)。
- **容器资源**需要完成[深度采集](#)后，才可以进行[迁移方案设计](#)。
- **中间件、数据库、存储**等资源可直接进行[迁移方案设计](#)。支持对AWS RDS（包括MySQL、MariaDB、Aurora、PGSQL、SQLServer、Oracle）和AWS DocumentDB数据库进行[深度采集](#)，以获取数据库更详细的信息。
- **对象存储资源**可以进行[深度采集](#)和[迁移方案配置](#)。

---结束

主机源端迁移准备度检查

请按照以下步骤对**主机资源**进行**迁移预检查（源端迁移准备度）**，确定源端主机是否满足迁移条件。

1. 在源端内网环境中[安装Edge工具](#)并完成注册。
2. 在[资源采集](#)页面的**资源列表**，单击主机或主机的资源数量，进入[资源列表](#)页面的**主机**页签。

资源列表

资源类型	资源数量
主机	138
容器	0
中间件	2
数据库	1
大数据	0
网络	496
存储	116

- 勾选待迁移的主机，单击列表上方的**关联到应用**按钮，弹出关联到应用窗口。
 - 已提前**创建应用**，在下拉列表中选择要关联的应用，单击“**确定**”。
 - 未提前创建应用，单击“**新建应用**”，输入自定义的**应用名称**和**描述**，选择**业务场景**、**使用场景**和**目的端区域**，单击“**创建**”按钮，应用创建成功。选择创建的应用后，单击“**确定**”。
- 在列表上方选择“**目标场景>主机迁移**”，此时**源端迁移准备度**列会出现“**配置**”按钮。
- 单击**源端迁移准备度**列的“**配置**”，弹出**配置**窗口。
- 根据**表2-13**，配置参数。

表 2-13 迁移准备度参数配置说明

参数	配置说明
类型	根据实际的主机操作系统类型选择。
采集设备	选择在源端环境安装的Edge工具。
接入IP	选择主机接入IP，可以为公网IP，也可以是私有IP；迁移预检查通过后，该IP会作为后续主机迁移工作流的资源接入IP。
端口	主机开放端口。 <ul style="list-style-type: none"> Windows主机默认为5985端口，无法修改。 Linux主机默认22端口，可根据实际情况修改。
凭证	选择主机凭证，如未提前在Edge上 添加源端主机凭证 ，请前往Edge控制台添加，并同步至迁移中心。 须知 为确保采集过程具备充分的系统访问权限，从而能够获取到必要的信息和数据。对主机深度采集的凭证要求如下： <ul style="list-style-type: none"> 对Linux主机进行深度采集时，请添加Root账号和密码作为采集凭证。 对Windows主机进行深度采集时，请添加Administrator账号和密码作为采集凭证。

- 配置完成后，单击“**确认**”按钮，系统会校验配置信息的正确性并开始源端迁移准备度检查，当**源端迁移准备度**列状态会变为**已就绪**时，代表该主机已满足迁移条件，可以进行下一阶段**迁移方案设计**或**迁移方案配置**。

主机深度采集

请按照以下步骤对**主机资源**进行深度采集。

1. 在源端内网环境中**安装Edge工具**并与迁移中心连接。
2. 在**资源采集**页面的**资源列表**，单击**主机资源**的**资源数量**，进入**资源列表**页面的**主机**页签。
3. 在需要进行深度采集的**主机采集设备**列，单击“**配置**”，弹出**配置**窗口。根据**表 2-14**，配置参数。

表 2-14 深度采集参数配置说明

参数	配置说明
类型	根据实际的主机操作系统类型选择。
采集设备	选择在源端环境安装的Edge工具。
接入IP	选择主机接入IP，可以为公网IP，也可以是私有IP；迁移预检查通过后，该IP会作为后续主机迁移工作流的资源接入IP。
端口	主机开放端口。 <ul style="list-style-type: none"> • Windows主机默认为5985端口，无法修改。 • Linux主机默认22端口，可根据实际情况修改。
凭证	选择主机凭证，如未提前在Edge上 添加源端主机凭证 ，请前往Edge控制台添加，并同步至迁移中心。 须知 为确保采集过程具备充分的系统访问权限，从而能够获取到必要的信息和数据。对主机深度采集的凭证要求如下： <ul style="list-style-type: none"> • 对Linux主机进行深度采集时，请添加Root账号和密码作为采集凭证。 • 对Windows主机进行深度采集时，请添加Administrator账号和密码作为采集凭证。

4. 单击“**确认**”，采集设备和凭证都绑定成功后，系统会自动开始深度采集。当深度采集列的状态为“**已采集**”时，代表采集完成。可以进行下一阶段**迁移方案设计**或**迁移方案配置**。

容器深度采集

请按照以下步骤对**容器资源**进行深度采集。

1. 在源端内网环境中**安装Edge工具**并与迁移中心连接。
2. 在**资源采集**页面的**资源列表**，单击**容器资源**的**资源数量**，进入**资源列表**页面的**容器**页签。
3. 在需要进行深度采集的**容器采集设备**列，单击“**绑定**”，弹出**绑定采集设备**窗口。
如果需要批量绑定，可以同时勾选多个资源，单击列表上方的“**绑定采集设备**”，弹出**绑定采集设备**窗口。
4. 选择源端安装的Edge工具，单击“**确定**”。绑定成功后，资源的采集设备列状态会变为已绑定。

5. 采集设备绑定成功后，还需要为资源绑定凭证，单击资源凭证列的“**绑定**”，弹出**绑定凭证**窗口。
6. 选择资源凭证，如未提前在Edge上**添加源端资源凭证**，请前往Edge控制台添加，并同步至迁移中心。
7. 单击“**确定**”，系统会自动检查凭证绑定状态。当深度采集列状态为**就绪**时，单击深度采集列的“**采集**”进行深度采集。首次深度采集完成后，可以单击采集状态列的“**重新采集**”按钮，进行多次深度采集。采集完成后，单击资源名称可以查看采集到的容器详情。

对象存储深度采集

通过深度采集获取对象存储资源的详细信息，以提高迁移集群规格评估的准确性。请按照以下步骤进行**对象存储资源**深度采集。

1. 在源端内网环境中**安装Edge工具**并与迁移中心连接。
2. 在**资源采集**页面的**资源列表**，单击**存储资源**的**资源数量**，进入**资源列表**页面的**存储**页签。
3. 在对象存储资源列表中，单击**深度采集配置**列的“**配置**”，弹出**配置**窗口。
4. 选择源端安装的Edge工具和相应的资源凭证，单击“**确认**”完成配置。如未提前在Edge上**添加源端资源凭证**，请前往Edge控制台添加，并同步至迁移中心。

须知

在对Azure云平台对象存储资源进行深度采集时，需要使用“**存储账户**”和“**密钥**”作为采集凭证，凭证的获取方法请参见**如何获取Azure对象存储深度采集所需凭证**。

5. 配置完成后，单击操作列的“**添加前缀**”，弹出**添加前缀**窗口。
6. 输入指定的资源目录路径进行采集，不输入则默认采集全桶资源，单击“**确定**”保存前缀设置。
7. 完成深度采集配置和添加前缀后，单击操作列的“**深度采集**”，系统开始进行对象存储资源的深度采集。支持进行多次深度采集。当采集状态为**采集完成**后，单击资源名称，可以查看采集到的信息。

数据库深度采集

当前支持对AWS RDS（包括MySQL、MariaDB、Aurora、Postgre SQL、SQL Server、Oracle）和AWS DocumentDB数据库进行深度采集，以获取包括数据库版本、引擎、服务器字符集、平均事务每秒(Transaction Per Second, TPS)和查询每秒(Query Per Second, QPS)等关键性能指标在内的详细信息。不同的数据库类型，采集的信息详情可能有所差异。

1. 在源端内网环境中或在可以访问到源端数据库的网络环境中**安装Edge工具**并与迁移中心连接。
2. 在**资源采集**页面的**资源列表**，单击**数据库资源**的**资源数量**，进入**资源列表**页面的**数据库**页签。
3. 在数据库资源列表中，通过**平台类型**过滤出采集到的所有AWS数据库资源，在支持深度采集的数据库类型的**采集设备**列，单击“**绑定**”，弹出**绑定采集设备**窗口。

如果需要批量绑定，可以同时勾选多个资源，单击页面右上角的“**绑定采集设备**”，弹出**绑定采集设备**窗口。

4. 选择源端安装的Edge工具，以及网络接入方式（如果采集的资源与安装Edge的主机处于同一VPC内，可以选择私网接入，否则需要选择公网接入），单击“**确定**”。绑定成功后，资源的采集设备列状态会变为已绑定。
5. 采集设备绑定成功后，还需要为资源绑定凭证，单击资源**凭证**列的“**绑定**”，弹出**绑定凭证**窗口。
6. 选择数据库资源凭证，如未提前在Edge上**添加源端资源凭证**，请前往Edge控制台添加，并同步至迁移中心。
7. 单击“**确定**”，系统会自动检查凭证绑定状态。当**深度采集**列的状态为**就绪**时，单击“**采集**”开始深度采集。首次深度采集完成/采集失败后，可以单击深度采集列的“**重新采集**”，进行多次深度采集。
8. 深度采集完成后，单击数据库名称，进入**数据库详情**页面，在**数据库详情**区域，可以查看采集到的详细信息。

查看发现的源端资源列表和资源详情

发现任务仅能获取源端资源的基本信息，更多详细信息需要通过深度采集获取。

1. 在**资源采集**页面的“**资源列表**”，可以查看发现的各类型资源数量和未关联应用的资源数量。

图 2-21 查看资源列表

资源类型	资源数量	未关联应用资源数量
主机	138	133 查看
容器	0	0
中间件	2 ▼	2 ▼
数据库	1	1 查看
大数据	0	0
网络	495 ▼	495 ▼
其他	118 ▼	118 ▼

2. 单击资源类型或资源数量，进入对应的资源列表。
3. 单击资源名称，可以查看该资源的基本信息。

查看任务状态和详情

步骤1 在**资源采集**页面，单击**任务总数**后的“**查看**”，进入任务列表页面。

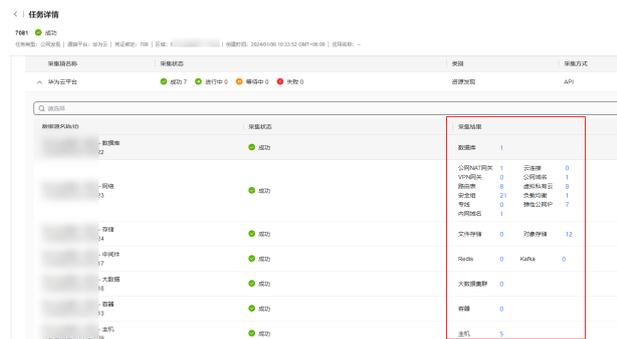
图 2-22 查看任务列表



步骤2 单击采集任务操作列的“查看”，进入任务详情页面，可以查看到该任务包含的采集项。

步骤3 单击采集项前的  按钮，可以查看该采集项包含的采集数据源以及各资源类型的采集结果。

图 2-23 采集结果



如果有采集失败的数据源，可以将鼠标放到数据源的采集状态上，查看失败原因。



步骤4 在采集结果列，单击资源类型后的数量，可以跳转到对应的资源类型列表页面。

---结束

2.4.2.2 内网发现与采集

在源端内网环境中部署Edge工具，通过网段扫描或VMware扫描的方式，发现和采集源端主机资源。

使用须知

- 目前支持采集的VMware版本包括：**VMware 5.0~7.0**

- VMware扫描和网段扫描两种方法都采用私有IP结合Edge ID的方式来标识内网中的主机资源。这意味着，一旦完成一次资源采集，如果被采集的主机私有IP发生变更，系统将在下一次采集时将其识别为一个新的主机资源，从而导致主机资源数量增加。为了避免这种情况，建议在迁移过程完成之前，不要更改源端主机的私有IP地址。

前提条件

- 已在源端内网环境[安装Edge工具](#)，并与迁移中心连接。
- 已在Edge上[添加源端主机凭证](#)。

须知

为确保采集过程具备充分的系统访问权限，从而能够获取到必要的信息和数据。对主机深度采集的凭证要求如下：

- 对Linux主机进行深度采集时，请添加Root账号和密码作为采集凭证。
- 对Windows主机进行深度采集时，请添加Administrator账号和密码作为采集凭证。

创建内网采集任务

步骤1 登录[迁移中心管理控制台](#)。

步骤2 单击左侧导航栏的“[调研>资源采集](#)”，进入[资源采集](#)页面。在页面左上角的[当前项目](#)下拉列表中选择[迁移项目](#)。

步骤3 单击任务框中的“[发现>内网发现](#)”，进入[创建内网发现任务](#)页面。

图 2-24 内网发现



步骤4 根据[表2-15](#)，配置内网发现任务参数。

表 2-15 内网发现任务参数说明

参数	说明
任务名称	自定义发现任务名称。
任务说明	输入发现任务说明。
设备	选择在源端内网环境中安装的Edge工具。

步骤5 根据实际情况，选择使用网段扫描或VMware扫描发现主机资源。

- 选择网段扫描，根据表2-16，配置网段扫描参数。

表 2-16 网段扫描参数说明

参数	说明
协议	选择通信协议TCP或ICMP。
网段	输入IP地址段，需在以下范围内： - 10.0.0.0 ~10.255.255.255 - 172.16.0.0 ~172.31.255.255 - 192.168.0.0 ~ 192.168.255.255
Linux	填写Linux系统类型主机需要扫描的端口。仅TCP协议需要填写，若不需要扫描Linux类型主机，可将端口设为0。
Windows	填写Windows系统类型主机需要扫描的端口。仅TCP协议需要填写，若不需要扫描Windows类型主机，可将端口设为0。

- 选择VMware扫描，在连接地址填写vCenter的IP地址，并选择vCenter的凭证。如未提前添加凭证，请前往迁移工具Edge添加私有云凭证。凭证中的用户名、密码为vCenter的登录用户名和密码。

步骤6 单击“确认”，内网发现任务创建完成，系统开始主机资源发现。

在资源采集页面，单击任务总数后的“查看”，进入任务列表页面，查看任务状态。

步骤7 内网发现任务状态为成功后，即可进行主机深度采集。

- 在资源采集页面的资源列表，单击主机资源的资源数量，进入资源列表页面的主机页签。
- 单击主机采集设备列表或凭证列的“配置”，弹出配置窗口。
- 根据表2-17，配置参数。

表 2-17 深度采集参数配置说明

参数	配置说明
类型	根据实际的主机操作系统类型选择。
采集设备	选择在源端环境安装的Edge工具。

参数	配置说明
接入IP	选择主机接入IP，可以为公网IP，也可以是私有IP；迁移预检查通过后，该IP会作为后续主机迁移工作流的资源接入IP。
端口	主机开放端口。 - Windows主机默认为5985端口，无法修改。 - Linux主机默认22端口，可根据实际情况修改。
凭证	选择主机凭证，如未提前在Edge上 添加源端主机凭证 ，请前往Edge控制台添加，并同步至迁移中心。

- 单击“**确认**”，采集设备和凭证都绑定成功后，系统会自动开始深度采集。当深度采集列的状态为“**已采集**”时，代表采集完成。可以进行下一阶段[迁移方案设计](#)或[迁移方案配置](#)。

图 2-25 采集完成



----结束

2.4.2.3 导入阿里云资源清单

从阿里云导出资源清单（CSV格式），再将其导入MgC，系统会自动解析这些文件，识别出相应的资源，并在资源列表展示资源的详细信息。此外，利用解析后的信息，您还可以使用迁移中心的规格评估功能，为您推荐与阿里云资源相匹配的华为云资源规格。

支持导入的阿里云资源类型包括：

- 云服务器 ECS实例和云盘
- 对象存储 OSS桶
- 云数据库 RDS MySQL实例
- 云数据库 Redis实例

导出阿里云 ECS 实例和云盘

步骤1 登录阿里云ECS控制台。

步骤2 在概览页面，单击“**导出数据>实例**”，导出所有实例资源清单。



步骤3 单击“导出数据>云盘”，导出所有云盘资源清单。

步骤4 打开实例资源清单和云盘资源清单，根据实际需求，调整清单内的资源信息。请确保清单中包含了以下必要信息：

清单类型	必选字段
实例清单	实例ID、操作系统、实例规格、CPU、内存
云盘清单	实例Id、云盘Id、容量（GiB）、云盘属性

----结束

导出阿里云 OSS 桶

步骤1 登录阿里云OSS控制台。

步骤2 在左侧导航栏选择“Bucket 列表”，进入Bucket 列表页面。

步骤3 单击列表右上角的  图标，弹出导出Bucket列表窗口。



步骤4 勾选需要导出的信息，单击“确定”，导出所有Bucket资源清单。

----结束

导出阿里云 RDS MySQL 实例

步骤1 登录阿里云RDS控制台。

步骤2 在左侧导航栏选择“实例列表”，进入实例列表页面。在页面上方选择地域。

步骤3 通过数据库类型列过滤出所有MySQL类型的实例，单击列表右上角的  图标。



步骤4 在弹出的对话框中，勾选需要导出的信息。单击“确定”。系统会将实例信息输出至CSV格式的文件并自动下载。

步骤5 由于从阿里云导出的CSV文件内容格式不满足MgC解析标准，因此需要进行如下修改：

1. 使用文件编辑器打开下载的资源清单CSV文件，在表头行搜索“PrivatePort(A network port,)”。
2. 将该表头名中的英文逗号删除，然后保存文件。正确的表头名应该为“PrivatePort(A network port)”。

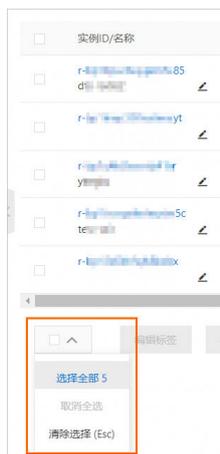
----结束

导出阿里云 Redis 实例

步骤1 登录阿里云云数据库 Redis 版控制台。

步骤2 在左侧导航栏选择“实例列表”，进入实例列表页面。在页面上方选择地域。

步骤3 在实例列表页面左下方，展开 ，单击“选择全部”。



步骤4 单击列表右上角的  图标，在弹出的对话框中，勾选需要导出的信息。



步骤5 单击“确定”。系统会将实例信息输出至CSV格式的文件并自动下载。

步骤6 由于从阿里云导出的CSV文件内容格式不满足MgC解析标准，因此需要进行如下修改：

1. 使用文件编辑器打开下载的资源清单CSV文件，
2. 在表头行的末尾，添加一个英文逗号，然后保存文件。

例如，原始表头为：

```
ID,Name,Quantity
```

修改后的表头应该是：

```
ID,Name,Quantity,
```

----结束

导入阿里云资源清单

步骤1 登录[迁移中心管理控制台](#)。

步骤2 单击左侧导航栏的“调研>资源采集”，进入资源采集页面。在页面左上角的当前项目下拉列表中选择[迁移项目](#)。

步骤3 单击任务框中的“导入>阿里云资源清单”，弹出导入阿里云资源清单窗口。

图 2-26 导入阿里云资源清单



须知

仅支持上传UTF-8编码且为中文表头的CSV文件，单个文件大小不能超过15MB。将CSV文件编码转换为UTF-8的方法请参考[转换CSV文件编码为UTF-8](#)。

步骤4 在资源类型下拉列表中，选择要导入的资源类型，并上传对应的资源清单。

导入阿里云资源清单

默认采用覆盖策略：若检测到相同实例ID的资源，新资源信息将自动覆盖旧资源信息。同时，新导入的关联和依赖关系将与原有关系整合更新。

* 资源类型 [如何获取资源清单](#)

* 资源清单
 仅支持上传UTF-8编码且为中文表头的CSV文件，且单个文件大小不能超过50MB。
[UTF-8编码转换方法](#)

* 云盘资源清单
 仅支持上传UTF-8编码且为中文表头的CSV文件，且单个文件大小不能超过50MB。

资源类型	操作步骤
主机	1. 单击资源清单后的“添加文件”，将导出的阿里云ECS实例资源清单上传。 2. 实例资源清单上传后，单击云盘资源清单后的“添加文件”，将导出的阿里云ECS云盘资源清单上传。
Redis	单击资源清单后的“添加文件”，将导出的阿里云Redis实例资源清单上传。
对象存储	单击资源清单后的“添加文件”，将导出的阿里云OSS桶资源清单上传。
数据库	单击资源清单后的“添加文件”，将导出的阿里云RDS MySQL实例资源清单上传。

步骤5 单击“**确认**”，资源清单上传成功后，在**资源列表**中选择导入的资源类型，进入对应的资源列表页面，可以查看导入的资源列表。

步骤6 （可选）将导入的阿里云资源**关联应用**，并使用**规格评估**功能，获取与阿里云资源相匹配的华为云资源规格。

----结束

2.4.2.4 导入 RVTools 资源

约束与限制

- **版本要求**
支持导入RVTools以下版本的导出结果：
 - 4.4.1
 - 4.4.2
 - 4.4.3
 - 4.4.4
 - 4.4.5
 - 4.5.0
 - 4.5.1
 - 4.6.1
- **文件格式要求**
导出结果的文件格式必须为Excel (.xlsx)。
- **文件大小、压缩比要求**
导入的文件大小不能超过100 MB，压缩比不能低于5%。

步骤一：导出 RVTools 资源

步骤1 启动RVTools应用程序，进入登录界面。

步骤2 在登录界面输入以下信息。

- 在IP地址栏中，输入vCenter服务器的IP地址。
- 在用户名栏中，填写用于连接vCenter的用户名。
- 在密码栏中，输入与用户名相对应的密码。

步骤3 单击“**登录**”按钮。登录成功后，在菜单栏单击“**file > Export all to Excel**”。

步骤4 选择保存路径，将资源导出到本地。

----结束

步骤二：将导出的文件上传至 OBS

步骤1 登录华为云**OBS控制台**，创建一个用于存储RVTools导出资源文件的**标准存储桶**。方法请参考**创建桶**。

步骤2 参考[上传对象](#)，将[导出的RVTools资源](#)文件上传至OBS标准存储桶。

----结束

步骤三：导入 RVTools 资源

步骤1 登录[迁移中心管理控制台](#)。

步骤2 单击左侧导航栏的“[调研>资源采集](#)”，进入[资源采集](#)页面。在页面左上角的[当前项目](#)下拉列表中选择[迁移项目](#)。

步骤3 单击任务框中的“[导入 > RVTools](#)”，弹出[创建RVTools导入任务](#)窗口。



说明

首次使用该功能时需要授权委托，单击RVTools导入委托窗口的“[同意授权](#)”即可。

步骤4 在文件的下拉列表中，找到[步骤二](#)存放RVTools导出资源文件的OBS桶，并选择导出的资源文件，单击“[确认](#)”按钮，上传至迁移中心。

步骤5 上传成功后，单击任务框中，任务总数后的“[查看](#)”，进入任务列表页面，查看任务状态。



- 如果任务状态为“[失败](#)”，单击操作列的“[查看](#)”，可以查看采集失败的数据源，将鼠标放到数据源的采集状态上，可以查看失败原因。常见失败原因的处理方法请参考[导入RVTools资源任务失败](#)。

处理完失败原因后，需要删除该采集任务，重新导入文件进行采集。





- 当任务状态为“成功”时，您可以查看导入的[资源详情](#)和[资源统计信息](#)。

----结束

查看主机资源详情

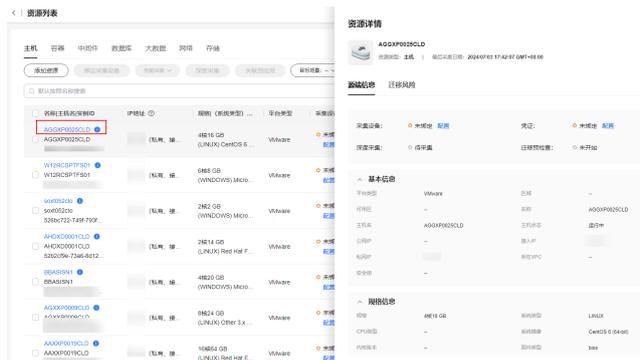
当RVTools导入任务状态为“成功”时，在资源列表页面，可以查看VMware采集的主机信息，包括：主机名称、实例ID、IP地址、主机规格、操作系统类型以及平台类型等。

- 步骤1** 在资源采集页面的资源列表，单击主机资源类型或资源数量，进入资源列表页面的主机页签。

资源列表

资源类型	资源数量
主机	138
容器	0
中间件	2
数据库	1
大数据	0
网络	496
存储	116

- 步骤2** 在主机资源列表，可以查看导入的主机资源信息，单击资源名称，可以查看该主机的更多基本信息、规格信息等。



----结束

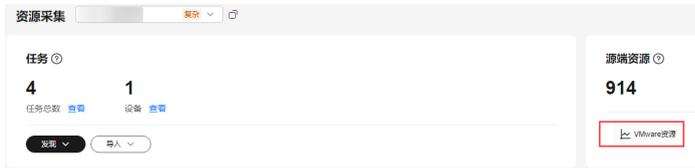
查看资源统计信息

当RVTools导入任务状态为“成功”时，在资源统计页面，可以查看VMware资源的统计信息，包括：

- 宿主机：CPU类型统计、宿主机个数统计、内存总量统计、已分配虚拟机内存、已分配占比、虚拟机总数统计、虚拟机运行总数统计、存储使用详情统计以及主机列表详情。

- 虚拟机：Windows操作系统个数统计、Linux系统个数统计、内存使用总容量、CPU使用核数、磁盘使用个数、磁盘容量占比、虚拟机数量、虚拟机列表、操作系统。

步骤1 在资源采集页面，单击源端资源框中的“VMware资源”，进入资源统计页面。



步骤2 分别选择VMware资源的“宿主机”页签和“虚拟机”页签，查看对应的统计信息。

----结束

2.4.3 手动添加资源

目前支持手动添加的资源类型包括：**主机**、**Oracle数据库**。通过在源端环境中安装Edge工具进行资源采集。

准备工作

- 在MgC控制台**创建项目**。
- 在源端内网环境**安装Edge工具**，并与**迁移中心连接**。
- 在Edge上**添加源端资源凭证**。

添加主机

步骤1 登录**迁移中心管理控制台**。

步骤2 单击左侧导航栏的“**调研>资源采集**”，进入**资源采集**页面。在页面左上角的当前项目下拉列表中选择**迁移项目**。

步骤3 在**资源列表**单击主机或主机的资源数量，进入**资源列表**页面的**主机**页签。

资源列表

资源类型	资源数量
主机	138
容器	0
中间件	2
数据库	1
大数据	0
网络	496
存储	116

步骤4 单击列表上方的“**添加资源**”按钮，弹出添加主机窗口。

图 2-27 添加主机



步骤5 根据表2-18，填写相应参数后，单击“确认”，系统会自动检查凭证绑定状态和进行信息采集。

表 2-18 添加主机参数说明

参数	说明
名称	用户自定义
采集设备	选择源端注册的Edge设备。
类型	主机实际系统类型
IP	填写主机IP。 如果采集的主机与安装Edge的主机处于同一VPC内，可以填写私有IP，否则需要填写公网IP。
端口	主机开放端口。 <ul style="list-style-type: none"> Windows主机默认为5985端口，无法修改。 Linux主机默认22端口，可根据实际情况修改。
凭证	选择源端主机凭证，如未提前在Edge上 添加源端主机凭证 ，请前往Edge控制台添加，并同步至迁移中心。

步骤6 添加成功后，在资源列表页面的主机页签，可以查看添加的主机资源。

----结束

添加数据库

步骤1 登录[迁移中心管理控制台](#)。

步骤2 单击左侧导航栏的“调研>资源采集”，进入资源采集页面。在页面左上角的当前项目下拉列表中选择[迁移项目](#)。

步骤3 在资源列表单击数据库或数据库的资源数量，进入资源列表页面的数据库页签。

资源类型	资源数量
主机	169
容器	10
中间件	10
数据库	14
大数据	0
网络	580
存储	135

步骤4 单击“添加资源”按钮，弹出添加数据库窗口。



步骤5 根据表2-19，填写相应参数。

×

添加数据库

* 名称

* 采集设备

* 类型

* IP
请确保采集设备能通过该地址连接到该资源。

* 端口

* 服务名称/SID

* 凭证

表 2-19 添加数据库参数配置说明

参数	说明
名称	用户自定义
采集设备	选择在源端注册并与迁移中心连接的Edge。
类型	选择“Oracle”。
IP	输入可以连接到源端Oracle数据库服务器的IP地址。
端口	源端Oracle 数据库监听端口。默认为1521。
服务名称/SID	源端数据库待采集的服务名。默认为ORCL。 您可以登录Oracle 数据库，使用如下命令，查看并确定需要采集的服务名称。 <code>SELECT instance_name FROM v\$instance;</code>
凭证	选择源端数据库凭证，如未提前在Edge上 添加源端数据库凭证 ，请前往Edge控制台添加，并同步至迁移中心。资源类型选择“数据库”，认证方式选择“用户名/密码”。

步骤6 参数配置完成后，单击“确认”，系统会自动进行深度采集。当深度采集列的状态变为**已采集**，代表采集完成。单击数据库名称，进入数据库详情页面，可以查看采集到的信息。



----结束

2.4.4 采集主机性能

迁移中心支持对主机资源进行性能数据采集，获取源端主机的CPU、内存负载、磁盘IOPS，内网出入流量，网络出入包速率，网络连接数等性能信息。在后续进行目的端规格评估时，可以根据采集到的源端主机实际性能负载推荐目的端主机CPU核数、内存大小等性能规格。

使用须知

- 开始采集后，默认每5分钟进行一次数据采集。
- 建议采集时长≥1小时，若采集时长不足，将无法获取7天和30天最大值以及平均值数据。
- 采集周期默认为7天，7天后会自动停止采集；采集周期内可以进行停止采集和重启采集操作；重启采集后，采集周期重新计算。
- 采集到的性能数据将保留180天，超期后会自动删除。
- 迁移工具Edge建议安装在独立主机上，避免与业务主机共用，以确保采集的稳定性和效率。
- Linux主机：迁移工具Edge最多同时采集1000台Linux主机。当同时采集1000台Linux主机时，安装Edge的主机规格最低要求为8U16G，且可使用内存不小于8G。
- Windows主机：迁移工具Edge最多同时采集500台Windows主机，当同时采集500台Windows主机时，安装Edge的主机规格最低要求为16U32G，且可使用内存不小于8G。

前提条件

- 已通过公网采集/内网采集/手动添加等方式完成主机资源采集。
- 已在源端内网环境安装Edge工具，并与迁移中心连接。
- 已在迁移工具 Edge上添加源端主机凭证。

操作步骤

步骤1 登录[迁移中心管理控制台](#)。

步骤2 单击左侧导航栏的“[调研>资源采集](#)”，进入[资源采集](#)页面。在页面左上角的当前项目下拉列表中选择[迁移项目](#)。

步骤3 在[资源列表](#)单击主机或主机的资源数量，进入[资源列表](#)页面的主机页签。

步骤4 为需要进行性能采集的主机配置[采集设备](#)和[凭证](#)。

- 如果性能采集列的“[开始](#)”按钮置灰，将鼠标移动到“[开始](#)”按钮处，在弹出的提示框中，单击“[去配置](#)”，弹出[配置](#)窗口。



- 如果性能采集列的“开始”按钮未置灰，代表已完成配置，可以跳过此步执行步骤7。

步骤5 根据表2-20，配置参数。

表 2-20 迁移准备度参数配置说明

参数	配置说明
类型	根据实际的主机操作系统类型选择。
采集设备	选择在源端环境安装的Edge工具。
IP	选择主机接入IP，可以为公网IP，也可以是私有IP；在需要使用代理连接主机的情况下，也可以自行输入接入IP。迁移预检查通过后，该IP会作为后续主机迁移工作流的资源接入IP。
端口	主机开放端口。 <ul style="list-style-type: none"> • Windows主机默认为5985端口，无法修改。 • Linux主机默认22端口，可根据实际情况修改。
凭证	选择主机凭证，如未提前在Edge上添加源端主机凭证，请前往Edge控制台添加，并同步至迁移中心。

步骤6 配置完成后，单击“确认”按钮，系统会校验配置信息的正确性。

步骤7 配置信息校验无误后，单击性能采集列的“开始”，开始主机性能采集，此时状态为“采集中”。

开始性能采集后，您还可以进行如下操作：

- **查看采集数据**
 - a. 单击主机名称，进入主机详情页面，在性能信息区域，可以查看采集到的主机性能数据。其中，7天和30天最大值为每小时汇总的定时任务，采集时长需超过1小时才能获取数据。
 - b. 单击操作列的“查看”，可以查看采集周期内每项数据的详细数值和曲线图。
- **停止采集**
单击“停止”，可以暂停性能数据采集。

----结束

2.4.5 关联应用

概述

通过应用将资源进行分组管理，便于后续进行规格评估和创建迁移工作流。

操作步骤

- 步骤1** 登录[迁移中心管理控制台](#)。
 - 步骤2** 单击左侧导航栏的“[调研>资源采集](#)”，进入[资源采集](#)页面。在页面左上角的当前项目下拉列表中选择[迁移项目](#)。
 - 步骤3** 在[资源列表](#)选择资源类型，进入对应的资源列表页面。
 - 步骤4** 勾选需要关联到同一应用的资源，单击页面左上角的“[关联到应用](#)”，弹出[关联到应用](#)窗口。
 - 步骤5** 在下拉列表中选择需要关联的应用。未提前创建应用，单击“[新建应用](#)”，输入自定义的[应用名称](#)和[描述](#)；根据实际需求，选择[业务场景](#)、[使用场景](#)以及[目的端区域](#)，单击“[创建](#)”按钮，应用创建成功。
 - 步骤6** 选择应用后，单击“[确定](#)”。在各资源的[已关联应用](#)列，查看资源所关联应用。
- 结束

2.5 应用管理

2.5.1 创建应用

您可以通过创建应用，将采集到的资源添加到应用中，对资源进行功能分组，便于后续进行规格评估和创建迁移工作流。

新建应用

- 步骤1** 登录[迁移中心管理控制台](#)。
- 步骤2** 单击左侧导航栏的“[调研>应用管理](#)”，进入[应用管理](#)页面。在页面左上角的当前项目下拉列表中选择[迁移项目](#)。
- 步骤3** 单击应用框中的“[新建应用](#)”按钮，弹出[新建应用](#)窗口。



- 步骤4** 输入自定义的[应用名称](#)和[描述](#)；选择[业务场景](#)和[使用场景](#)；[目的端区域](#)选择目的端所在区域，单击“[创建应用](#)”按钮，应用创建成功，进入[关联资源](#)页面。
 - 如果已经进行过资源采集，并且需要将采集到的资源关联到创建的应用，请勾选需要关联到应用的资源，单击“[关联到应用](#)”按钮，完成应用创建和资源关联。

- 如果还未进行资源采集或暂不需要关联资源，单击“**暂不关联资源**”按钮，完成应用创建。后续在**资源列表**中也可以进行关联操作。

----结束

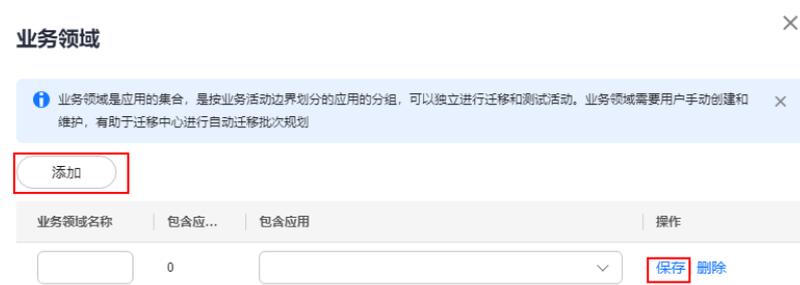
管理应用

您可以根据不同的业务领域将应用进行分类管理。

步骤1 在**应用管理**页面，单击**应用框**中业务领域的“**查看**”按钮，右侧弹出**业务领域**窗口。



步骤2 单击“**添加**”按钮，输入业务领域名称，选择该业务包含的应用，单击“**保存**”。



----结束

修改应用

仅简单项目支持。

步骤1 在应用列表，单击应用操作列的“**修改**”，弹出**修改应用**窗口。

步骤2 根据实际情况，修改应用参数后，单击“**确定**”，完成应用修改。

----结束

2.5.2 关联资源

本节介绍为应用关联资源的方法。通过应用将资源进行分组管理，为后续规格评估、创建迁移 workflow 作准备。

前提条件

已通过资源采集页面中**公网发现/内网发现/手动添加**等方式完成主机资源的发现。

操作步骤

- 步骤1** 登录[迁移中心管理控制台](#)。
- 步骤2** 单击左侧导航栏的“[调研>应用管理](#)”，进入[应用管理](#)页面。在页面左上角的[当前项目](#)下拉列表中选择[迁移项目](#)。
- 步骤3** 在[应用列表](#)，单击应用名称，进入[应用详情](#)页面。
- 步骤4** 在资源列表选择要添加的资源类型。以下步骤以关联主机资源为例。其它资源类型的关联方法与主机资源的方法相同。
- 步骤5** 在主机页签，单击列表上方的“[关联资源](#)”按钮，右侧弹出[关联资源](#)窗口。

图 2-28 关联资源



- 步骤6** 勾选需要关联到同一应用的主机资源，单击窗口右下角的“[关联到应用](#)”。在资源列表中即可看到关联到应用的主机资源。
- 步骤7** 应用完成资源关联后，单击页面右上角的“[规格评估](#)”按钮，进入应用评估页面。对应用内的资源进行规格评估以获取最合适的华为云对应资源规格。更多详细介绍请参考[规格评估](#)。

----结束

2.5.3 手动添加资源

支持在应用内手动添加[主机](#)和[Oracle数据库](#)两种资源类型。通过在源端环境中安装Edge工具进行资源采集。

准备工作

- 在MgC控制台[创建项目](#)。
- 在源端内网环境[安装Edge工具](#)，并与[迁移中心连接](#)。
- 在Edge上[添加源端资源凭证](#)。
- 已在应用管理页面[创建应用](#)。

添加主机

- 步骤1** 登录[迁移中心管理控制台](#)。
- 步骤2** 单击左侧导航栏的“[调研>应用管理](#)”，进入[应用管理](#)页面。在页面左上角的[当前项目](#)下拉列表中选择[迁移项目](#)。
- 步骤3** 在[应用列表](#)，单击应用名称，进入[应用详情](#)页面。
- 步骤4** 在资源列表的主机页签，单击列表上方的“[添加资源](#)”按钮，弹出添加主机窗口。

图 2-29 添加主机资源



步骤5 根据表2-21，填写相应参数后，单击“确认”，系统会自动检查凭证绑定状态和进行信息采集。

表 2-21 添加主机参数说明

参数	说明
名称	用户自定义
采集设备	选择源端注册的Edge设备。
类型	主机实际系统类型
IP	填写主机IP。 如果采集的主机与安装Edge的主机处于同一VPC内，可以填写私有IP，否则需要填写公网IP。
端口	主机开放端口。 <ul style="list-style-type: none"> Windows主机默认为5985端口，无法修改。 Linux主机默认22端口，可根据实际情况修改。
凭证	选择源端主机凭证，如未提前在Edge上 添加源端主机凭证 ，请前往Edge控制台添加，并同步至迁移中心。

步骤6 添加成功后，在资源列表页面的主机页签，可以查看添加的主机资源。

----结束

添加数据库

步骤1 登录[迁移中心管理控制台](#)。

步骤2 单击左侧导航栏的“调研>应用管理”，进入应用管理页面。在页面左上角的当前项目下拉列表中选择[迁移项目](#)。

步骤3 在应用列表，单击应用名称，进入应用详情页面。

步骤4 在资源列表的数据库页签，单击列表上方的“添加资源”按钮，弹出添加数据库窗口。

图 2-30 添加数据库资源



步骤5 根据表2-22，填写相应参数。

×

添加数据库

* 名称

* 采集设备

* 类型

* IP 请确保采集设备能通过该地址连接到该资源。

* 端口

* 服务名称/SID

* 凭证

表 2-22 添加数据库参数配置说明

参数	说明
名称	用户自定义
采集设备	选择在源端注册并与迁移中心连接的Edge。
类型	选择“Oracle”。
IP	输入可以连接到源端Oracle数据库服务器的IP地址。
端口	源端Oracle 数据库监听端口。默认为1521。
服务名称/SID	源端数据库待采集的服务名。默认为ORCL。 您可以登录Oracle 数据库，使用如下命令，查看并确定需要采集的服务名称。 <code>SELECT instance_name FROM v\$instance;</code>
凭证	选择源端数据库凭证，如未提前在Edge上 添加源端数据库凭证 ，请前往Edge控制台添加，并同步至迁移中心。资源类型选择“数据库”，认证方式选择“用户名/密码”。

步骤6 参数配置完成后，单击“确认”，系统会自动进行深度采集。当深度采集列的状态变为**已采集**，代表采集完成。单击数据库名称，进入数据库详情页面，可以查看采集到的信息。



----结束

2.5.4 采集主机性能

迁移中心支持对主机资源进行性能数据采集，获取源端主机的CPU、内存负载、磁盘IOPS，内网出入流量，网络出入包速率，网络连接数等性能信息。在后续进行目的端规格评估时，可以根据采集到的源端主机实际性能负载推荐目的端主机CPU核数、内存大小等性能规格。

本文介绍在应用详情内，对应用中的主机资源进行性能采集的操作方法。

使用须知

- 开始采集后，默认每5分钟进行一次数据采集。
- 建议采集时长 ≥ 1 小时，若采集时长不足，将无法获取7天和30天最大值以及平均值数据。
- 采集周期默认为7天，7天后会自动停止采集；采集周期内可以进行停止采集和重启采集操作；重启采集后，采集周期重新计算。
- 采集到的性能数据将保留180天，超期后会自动删除。
- 迁移工具Edge建议安装在独立主机上，避免与业务主机共用，以确保采集的稳定性和效率。
- Linux主机：迁移工具Edge最多同时采集1000台Linux主机。当同时采集1000台Linux主机时，安装Edge的主机规格最低要求为8U16G，且可使用内存不小于8G。
- Windows主机：迁移工具Edge最多同时采集500台Windows主机，当同时采集500台Windows主机时，安装Edge的主机规格最低要求为16U32G，且可使用内存不小于8G。

前提条件

- 已通过[关联资源](#)或[手动添加](#)等方式获取主机资源。
- 已在源端内网环境[安装Edge工具](#)，并与迁移中心连接。
- 已在迁移工具 Edge上[添加源端主机凭证](#)。

操作步骤

- 步骤1** 登录[迁移中心管理控制台](#)。
- 步骤2** 单击左侧导航栏的“[调研>应用管理](#)”，进入[应用管理](#)页面。在页面左上角的当前项目下拉列表中选择[迁移项目](#)。
- 步骤3** 在[应用列表](#)，单击应用名称，进入[应用详情](#)页面。
- 步骤4** 为需要进行性能采集的主机配置[采集设备](#)和[凭证](#)。
 - 如果性能采集列的“[开始](#)”按钮置灰，将鼠标移动到“[开始](#)”按钮处，在弹出的提示框中，单击“[去配置](#)”，弹出[配置](#)窗口。



- 如果性能采集列的“开始”按钮未置灰，代表已完成配置，可以跳过此步执行**步骤7**。

步骤5 根据**表2-23**，配置参数。

表 2-23 迁移准备度参数配置说明

参数	配置说明
类型	根据实际的主机操作系统类型选择。
采集设备	选择在源端环境安装的Edge工具。
IP	选择主机接入IP，可以为公网IP，也可以是私有IP；在需要使用代理连接主机的情况下，也可以自行输入接入IP。迁移预检查通过后，该IP会作为后续主机迁移工作流的资源接入IP。
端口	主机开放端口。 <ul style="list-style-type: none"> • Windows主机默认为5985端口，无法修改。 • Linux主机默认22端口，可根据实际情况修改。
凭证	选择主机凭证，如未提前在Edge上 添加源端主机凭证 ，请前往Edge控制台添加，并同步至迁移中心。

步骤6 配置完成后，单击“确认”按钮，系统会校验配置信息的正确性。

步骤7 配置信息校验无误后，单击性能采集列的“开始”，开始主机性能采集，此时状态为“采集中”。

开始性能采集后，您还可以进行如下操作：

- **查看采集数据**
 - a. 单击主机名称，进入**主机详情**页面，在**性能信息**区域，可以查看采集到的主机性能数据。其中，7天和30天最大值为每小时汇总的定时任务，采集时长需超过1小时才能获取数据。
 - b. 单击操作列的“查看”，可以查看采集周期内每项数据的详细数值和曲线图。
- **停止采集**
单击“停止”，可以暂停性能数据采集。

----结束

2.5.5 查看应用关联分析结果

迁移中心会整理源端资源和应用间的关联关系，并且绘制技术架构图和应用依赖图，为后续迁移计划的制定和目的端架构设计提供帮助。

前提条件

已导入应用关联文件或导入工具采集结果。

查看技术架构图

步骤1 在应用管理页面，单击应用框中的“技术架构图”，可以查看所有服务、微服务、资源的矩阵图。

步骤2 选择服务，单击鼠标右键，可以查看该服务的技术架构图、服务依赖图和包含的微服务列表。

----结束

查看应用列表

在应用管理页面的应用列表，可以查看项目下的所有应用。通过应用列表可以查看应用名称、应用类型、所属业务领域、包含微服务数量以及依赖应用和被依赖应用数量等信息。

- 单击包含微服务列的数量，可以查看该应用所包含的微服务列表。
- 单击依赖应用列的数量，可以查看该应用所依赖的服务列表。
- 单击被依赖应用列的数量，可以查看依赖该应用的应用列表。
- 单击操作列的“技术架构图”，可以查看该应用包含的微服务、资源所属层级以及微服务与资源之间的关联关系。单击微服务/资源节点，可以查看节点详细信息。
- 单击操作列的“应用依赖图”，可以查看该应用与依赖服务、被依赖服务之间的关联关系。单击服务节点，可以查看服务的详细信息和依赖服务、被依赖服务列表。

查看资源线索

步骤1 在应用管理页面，单击源端资源框中的的“线索”按钮，进入线索页面。

步骤2 选择资源线索页签，可以查看采集到的所有资源列表以及资源的详细信息。

- 单击操作列的“服务依赖图”，可以查看该资源与依赖服务、被依赖服务之间的关联关系。
- 单击操作列的“删除”，可以删除该资源。

----结束

查看微服务信息

步骤1 在应用管理页面，单击源端资源框中的的“线索”按钮，进入线索页面。

步骤2 选择微服务列表页签，可以查看采集到的所有微服务列表。

- 单击操作列的“技术架构图”，可以查看该微服务所属层级以及与资源之间的关联关系。单击微服务/资源节点，可以查看节点详细信息。
- 单击操作列的“更多>服务依赖图”，可以查看该微服务与依赖服务、被依赖服务之间的关联关系。单击微服务节点，可以查看微服务的详细信息和依赖服务、被依赖服务列表。

- 单击操作列的“**更多>删除**”，可以删除该微服务。

----结束

导出应用关联信息

在**应用管理**页面，单击应用框中的“**导出**”，可以导出接入层调用链、数据库依赖表、中间件依赖表、应用关联表，用于下线分析。

2.6 大数据血缘

2.6.1 部署迁移工具 Edge

使用迁移中心的大数据血缘功能前，需要先可以在可以访问源端元数据来源的环境中部署迁移工具Edge。

注意

迁移工具Edge不具备自动重启功能，严禁在任务执行期间重启Edge，以避免任务失败。

操作步骤

步骤1 安装Edge的方法请参考[下载并安装Edge](#)。

步骤2 安装成功后，将Edge与迁移中心进行连接，方法请参考[连接迁移中心](#)。

步骤3 Edge与迁移中心连接成功后，添加大数据相关凭证，方法请参考[添加资源凭证](#)。

----结束

2.6.2 采集元数据

2.6.2.1 创建元数据连接

创建元数据来源类型的连接，获取源端大数据数据库表清单。

目前支持的连接类型，参见[表2-24](#)。

表 2-24 支持的元数据连接类型

类别	连接类型
大数据组件	<ul style="list-style-type: none"> ● Hive Metastore ● Doris ● HBase ● ClickHouse ● Delta Lake (有元数据) ● Delta Lake (无元数据) ● Hudi (有元数据) ● Hudi (无元数据)
大数据云服务	<ul style="list-style-type: none"> ● MaxCompute ● 阿里云云数据库ClickHouse

须知

- 在创建Hive Metastore连接前，需要先获取连接Hive Metastore的IP地址和端口号。
- 在创建MaxCompute连接前，需要先获取MaxCompute服务所在地域的Endpoint信息。
- 在创建Doris连接前，需要先获取Doris集群部署的服务器IP地址和端口。
- 在创建HBase连接前，需要先获取HBase集群中至少一个ZooKeeper节点的IP地址和连接端口。

前提条件

已在源端元数据来源的环境中部署迁移工具Edge。

操作步骤

步骤1 登录[迁移中心](#)管理控制台。

步骤2 在左侧导航栏选择“[调研>大数据血缘](#)”，进入[大数据血缘](#)页面。在页面左上角的[当前项目](#)下拉列表中选择[迁移项目](#)。

步骤3 在[元数据采集](#)区域，单击“[创建连接](#)”，右侧弹出[创建连接](#)窗口。



步骤4 选择连接类型，单击“[下一步](#)”，进入[详细配置](#)页面。

步骤5 根据所选的连接类型，查看对应的参数配置说明，配置连接参数。

- [Hive Metastore连接参数配置说明](#)

- [Doris连接参数配置说明](#)
- [HBase连接参数配置说明](#)
- [ClickHouse组件参数配置说明](#)
- [Delta Lake（有元数据）连接参数配置说明](#)
- [Delta Lake（无元数据）连接参数配置说明](#)
- [Hudi（有元数据）连接参数配置说明](#)
- [Hudi（无元数据）连接参数配置说明](#)
- [MaxCompute连接参数配置说明](#)
- [阿里云云数据库ClickHouse连接参数配置说明](#)

表 2-25 Hive Metastore 连接参数配置说明

参数	配置说明
所属源	仅需在源端创建元数据连接。
连接名称	默认创建名称为“Hive-Metastore-4位随机数（字母和数字）”，也可以自定义名称。
迁移工具	选择在源端环境中 部署的迁移工具Edge 。
启用安全连接	选择是否启用安全连接。 <ul style="list-style-type: none"> • MetaStore在非安全集群下，可以不启用安全连接。 • MetaStore在安全集群下，需要启用安全连接并选择凭证。凭证的获取以及添加方法请参考添加资源凭证中的大数据-Hive Metastore。
Hive版本	选择源端Hive的实际版本。 注意 如果源端Hive为2.1.1版本，此处需要选择1.x版本。
MetaStore节点IP	填写连接MetaStore的IP地址。
MetaStore Thrift服务端口	填写连接MetaStore的端口，默认为9083。

表 2-26 Doris 连接参数配置说明

参数	配置说明
所属源	仅需在源端创建元数据连接。
连接名称	默认创建名称为“Doris-4位随机数（字母和数字）”，也可以自定义名称。
迁移工具	选择在源端环境中 部署的迁移工具Edge 。
Doris凭证	选择在Edge添加的Doris凭证。凭证的添加方式请参考 添加资源凭证 ，类型选择“大数据-Doris”。

参数	配置说明
数据库地址	填写连接到Doris集群部署的服务器IP地址。
数据库端口	填写连接到Doris集群部署的服务器的端口。默认为3306。
数据库名	填写Doris数据库名称。

表 2-27 HBase 连接参数配置说明

参数	配置说明
所属源	仅需在源端创建元数据连接。
连接名称	默认创建名称为“HBase-4位随机数（字母和数字）”，也可以自定义名称。
迁移工具	选择在源端环境中部署的迁移工具Edge。 注意 需要在安装Edge的主机上配置/etc/hosts后，HBase元数据连接才能连接成功。host文件的获取方法请参见 准备开发和运行环境 。
HBase凭证	选择在Edge添加的HBase凭证。凭证的添加方式请参考 添加资源凭证 ，类型选择“大数据-HBase”。
安全集群	选择是否部署在安全集群下。
Zookeeper节点IP	填写连接到Zookeeper的IP地址。通常是部署Zookeeper的服务器公网IP地址或内网IP地址。
Zookeeppeer连接端口	填写连接到Zookeeppeer的端口。
HBase 版本	选择HBase的版本。

表 2-28 ClickHouse 连接参数配置说明

参数	配置说明
所属源	仅需在源端创建元数据连接。
连接名称	默认创建名称为“ClickHouse-4位随机数（字母和数字）”，也可以自定义名称。
迁移工具	选择在源端环境中部署的迁移工具Edge。
ClickHouse凭证（可选）	选择在Edge添加的ClickHouse凭证。凭证的添加方法请参考 添加资源凭证 中的大数据-ClickHouse。
安全集群	选择是否部署在安全集群下。
ClickHouse Server IP	填写连接到ClickHouse服务器的IP地址。通常是运行ClickHouse的主机IP地址。

参数	配置说明
HTTP端口	非安全集群下，与ClickHouse 服务器通信的 HTTP 端口。默认为8123。
HTTP SSL/TLS端口	安全集群下，与ClickHouse 服务器通信的HTTPS 端口。

表 2-29 Delta Lake（有元数据）连接参数配置说明

参数	配置说明
所属源	仅需在源端创建元数据连接。
连接名称	默认创建名称为“Delta-Lake-有元数据-4位随机数（字母和数字）”，也可以自定义名称。
迁移工具	选择在源端环境中部署的迁移工具Edge。
执行机凭证	选择执行机的登录凭证。凭证的添加方法请参考 添加资源凭证 中的大数据-执行机。
执行机IP	填写连接执行机的IP地址。
执行机连接端口	填写连接执行机的端口。
Spark客户端目录	填写Spark客户端的安装目录路径。
环境变量地址	填写环境变量文件（配置文件）的绝对路径，例如：/opt/bigdata/client/bigdata_env。
SQL文件地址	填写用于存放执行作业时生成的SQL文件的文件夹地址。您需要拥有该文件夹的读写权限。 须知 迁移完成后，您需要手动清理该路径下生成的文件夹，以释放存储空间。

表 2-30 Delta Lake（无元数据）连接参数配置说明

参数	配置说明
所属源	仅需在源端创建元数据连接。
连接名称	默认创建名称为“Delta-Lake-无元数据-4位随机数（字母和数字）”，也可以自定义名称。
迁移工具	选择在源端环境中部署的迁移工具Edge。
执行机凭证	选择执行机的登录凭证。凭证的添加方法请参考 添加资源凭证 中的大数据-执行机。
执行机IP	填写连接执行机的IP地址。
执行机连接端口	填写连接执行机的端口。

参数	配置说明
Spark客户端目录	填写Spark客户端的安装目录路径。
环境变量地址	填写环境变量文件（配置文件）的绝对路径，例如：/opt/bigdata/client/bigdata_env。
SQL文件地址	填写用于存放执行作业时生成的SQL文件的文件夹地址。您需要拥有该文件夹的读写权限。 须知 迁移完成后，您需要手动清理该路径下生成的文件夹，以释放存储空间。

表 2-31 Hudi（有元数据）连接参数配置说明

参数	配置说明
所属源	仅需在源端创建元数据连接。
连接名称	默认创建名称为“Hudi-有元数据-4位随机数（字母和数字）”，也可以自定义名称。
迁移工具	选择在源端环境中部署的迁移工具Edge。
执行机凭证	选择执行机的登录凭证。凭证的添加方法请参考添加资源凭证中的大数据-执行机。
执行机IP	填写连接执行机的IP地址。
执行机连接端口	填写连接执行机的端口。
Spark客户端目录	填写Spark客户端的安装目录路径。
环境变量地址	填写环境变量文件（配置文件）的绝对路径，例如：/opt/bigdata/client/bigdata_env。
SQL文件地址	填写用于存放执行作业时生成的SQL文件的文件夹地址。您需要拥有该文件夹的读写权限。 须知 迁移完成后，您需要手动清理该路径下生成的文件夹，以释放存储空间。

表 2-32 Hudi（无元数据）连接参数配置说明

参数	配置说明
所属源	仅需在源端创建元数据连接。
连接名称	默认创建名称为“Hudi-无元数据-4位随机数（字母和数字）”，也可以自定义名称。
迁移工具	选择在源端环境中部署的迁移工具Edge。

参数	配置说明
执行机凭证	选择执行机的登录凭证。凭证的添加方法请参考 添加资源凭证 中的大数据-执行机。
执行机IP	填写连接执行机的IP地址。
执行机连接端口	填写连接执行机的端口。
Spark客户端目录	填写Spark客户端的安装目录路径。
环境变量地址	填写环境变量文件（配置文件）的绝对路径，例如：/opt/bigdata/client/bigdata_env。
SQL文件地址	填写用于存放执行作业时生成的SQL文件的文件夹地址。您需要拥有该文件夹的读写权限。 须知 迁移完成后，您需要手动清理该路径下生成的文件夹，以释放存储空间。

表 2-33 MaxCompute 连接参数配置说明

参数	配置说明
所属源	仅需在源端创建元数据连接。
连接名称	默认创建名称为“MaxCompute-4位随机数（字母和数字）”，也可以自定义名称。
迁移工具	选择在源端环境中 部署的迁移工具Edge 。
阿里云凭证	选择源端阿里云账号的凭证。凭证的添加方式请参考 添加资源凭证 ，类型选择“大数据-MaxCompute”。
项目名称	输入需要采集的MaxCompute项目名称。项目名称可在MaxCompute侧控制台获取。
Endpoint	输入MaxCompute服务所在地域的Endpoint信息。各地域Endpoint信息请参考 Endpoint对照表 。

表 2-34 阿里云云数据库 ClickHouse 连接参数配置说明

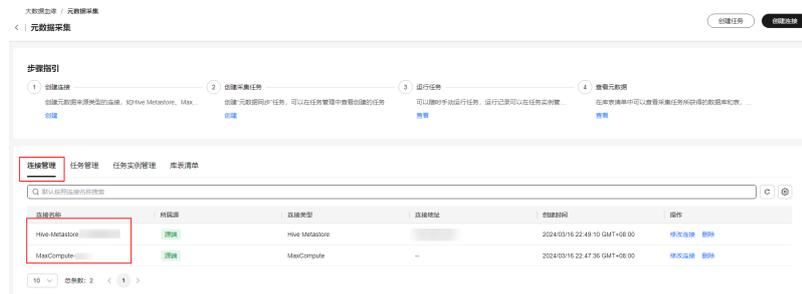
参数	配置说明
所属源	仅需在源端创建元数据连接。
连接名称	默认创建名称为“阿里云云数据库ClickHouse-4位随机数（字母和数字）”，也可以自定义名称。
迁移工具	选择在源端环境中 部署的迁移工具Edge 。
ClickHouse凭证	选择在Edge添加的源端阿里云云数据库ClickHouse凭证。凭证的添加方法请参考 添加资源凭证 中的大数据-ClickHouse。

参数	配置说明
数据库链接	填写连接源端ClickHouse集群的外网地址。可以在 集群信息 页面，查看外网地址。

步骤6 配置完成后，单击“**测试连接**”按钮，系统会校验配置信息并与源端建立连接，测试通过代表连接成功。

步骤7 测试连接通过后，单击“**确认**”按钮，创建元数据连接完成。

步骤8 单击“**元数据连接**”，进入元数据采集页面。在**连接管理**页签，可以查看已创建的元数据连接列表以及基本信息。单击操作列的“**修改连接**”，可以修改连接参数配置。



----结束

2.6.2.2 创建并运行元数据同步任务

利用元数据连接，创建元数据同步任务，将源端数据库的库、表、字段信息同步至迁移中心。

支持创建元数据同步任务的连接类型如下：

- **Hive Metastore**连接
- **Doris**连接
- **HBase**连接
- **ClickHouse**连接
- **Hudi**（有元数据）连接
- **Hudi**（无元数据）连接
- **MaxCompute**连接
- **阿里云云数据库ClickHouse**连接

您可以通过手动添加、导入模板等方式添加元数据库表信息，方法请参见[查看元数据](#)。

前提条件

已[创建元数据连接](#)。

操作步骤

步骤1 登录**迁移中心**管理控制台。

步骤2 在左侧导航栏选择“**调研>大数据血缘**”，进入**大数据血缘**页面。在页面左上角的**当前项目**下拉列表中选择**迁移项目**。

步骤3 在**元数据采集**区域，单击“**创建任务 > 元数据同步**”按钮，右侧弹出**创建任务-元数据同步**窗口。



步骤4 根据**表2-35**，配置元数据同步任务参数。

表 2-35 元数据同步任务参数配置说明

参数	配置说明
任务名称	默认创建名称为“元数据同步任务-4位随机数（字母和数字）”，也可以自定义名称。
元数据连接	根据实际需求，选择 已创建的元数据连接 类型。
数据库	填写需要进行元数据同步的数据库名称。 须知 元数据连接为 MaxCompute 类型时，该参数为必填参数，不能为空。其他元数据连接类型，该参数为可选参数，不填则代表同步全部数据库。
并发线程数	设置采集线程并发数。默认为3， $1 \leq \text{取值范围} \leq 10$ 。增加线程并发数，可以提高同步效率，但会占用更多源端连接和迁移工具（Edge）所安装主机的计算资源。

步骤5 参数配置完成后，单击“**确认**”按钮，创建元数据同步任务完成。

步骤6 单击“**采集任务**”，进入**元数据采集**页面。在**任务管理**页签，可以查看已创建的元数据同步任务列表以及参数配置信息。单击操作列的“**更多>修改**”，可以修改任务参数配置。



步骤7 单击操作列的“**运行任务**”，即可触发一次任务执行。每执行一次任务，会生成一条任务实例。

步骤8 单击操作列的“**查看实例**”，进入**任务实例管理**页签。可以查看该任务的运行记录，以及每条任务实例的状态和采集结果。当元数据同步任务的实例状态为“**已完成**”且采集结果有数值时，在**库表清单**页签可以查看采集到的元数据列表。



步骤9 元数据采集完成后，在元数据列表单击血缘列的“去采集”，可以创建血缘采集任务。



----结束

2.6.2.3 创建并运行数据湖元数据同步任务

利用创建的Delta Lake元数据连接，将源端数据库的库、表、字段信息同步至迁移中心。仅Delta Lake（有元数据）支持创建数据湖元数据同步任务。

Delta Lake（无元数据）只能通过手动添加、导入模板等方式添加元数据库表信息，方法请参见查看元数据。

前提条件

已创建Delta Lake（有元数据）元数据连接。

操作步骤

步骤1 登录迁移中心管理控制台。

步骤2 在左侧导航栏选择“调研>大数据血缘”，进入大数据血缘页面。在页面左上角的当前项目下拉列表中选择迁移项目。

步骤3 在元数据采集区域，单击“创建任务 > 数据湖元数据同步”按钮，右侧弹出创建任务-数据湖元数据同步窗口。



步骤4 根据表2-36，配置数据湖元数据同步任务参数。

表 2-36 元数据同步任务参数配置说明

参数	配置说明
任务名称	默认创建名称为“数据湖元数据同步任务-4位随机数（字母和数字）”，也可以自定义名称。
元数据连接	选择已创建的Delta-Lake（有元数据）元数据连接。
数据库	填写需要进行元数据同步的数据库名称。

参数	配置说明
并发线程数	设置采集线程并发数。默认为3，1≤取值范围≤10。增加线程并发数，可以提高同步效率，但会占用更多源端连接和迁移工具（Edge）所安装主机的计算资源。
自定义参数	<p>可以通过自定义参数设置同步的特定表或分区、过滤条件等。</p> <ul style="list-style-type: none"> 如果需要同步的是阿里云 EMR环境中的元数据，请添加如下参数： <ul style="list-style-type: none"> 参数名：conf 参数值：spark.sql.catalogImplementation=hive 如果源端是阿里云 EMR Delta 2.2版本并且是通过Delta 2.3版本依赖访问时，请添加如下参数： <ul style="list-style-type: none"> 参数名：master 参数值：local 如果源端是阿里云 EMR Delta 2.1.0版本并且配置了Spark 2.4.8来处理Delta Lake数据，请添加如下参数： <ul style="list-style-type: none"> 参数名：mgc.delta.spark.version 参数值：2 如果源端是阿里云 EMR并且配置了Spark 3来处理Delta Lake数据，请添加如下参数： <ul style="list-style-type: none"> 参数名：jars 参数值：'/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-core_2.12-*.jar,/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-storage-*.jar' <p>注意 参数值需要根据实际环境目录和Delta Lake 的版本进行替换。</p>

步骤5 参数配置完成后，单击“确认”按钮，创建数据湖元数据同步任务完成。

步骤6 单击“采集任务”，进入元数据采集页面。在任务管理页签，可以查看已创建的数据湖元数据同步任务列表以及参数配置信息。单击操作列的“更多>修改”，可以修改任务参数配置。



步骤7 单击操作列的“运行任务”，即可触发一次任务执行。每执行一次任务，会生成一条任务实例。

步骤8 单击操作列的“查看实例”，进入任务实例管理页签。可以查看该任务的运行记录，以及每条任务实例的状态和采集结果。当数据湖元数据同步任务的实例状态为“已完成”且采集结果有数值时，在库表清单页签可以查看采集到的元数据列表。



步骤9 数据湖元数据采集完成后，在元数据列表单击血缘列的“去采集”，可以[创建血缘采集任务](#)。



----结束

2.6.2.4 查看元数据

您可以通过创建采集任务、手动添加、导入模板等方式获取元数据库表清单。

通过采集任务获取元数据

创建采集任务的方法请参考[创建并运行元数据同步任务](#)。当元数据同步任务的实例状态为“已完成”且采集结果有数值时，在[库表清单](#)页签可以查看采集到的元数据列表。

手动添加数据表

步骤1 登录[迁移中心](#)管理控制台。

步骤2 在左侧导航栏选择“[调研>大数据血缘](#)”，进入[大数据血缘](#)页面。在页面左上角的当前项目下拉列表中选择[迁移项目](#)。

步骤3 在元数据采集区域，单击“[查看库表清单](#)”，进入元数据采集页面的[库表清单](#)页签。

步骤4 单击“[新增数据表](#)”按钮，弹出[新增数据表](#)窗口。

步骤5 根据[表2-37](#)，配置参数。

表 2-37 新增数据表参数说明

参数	说明
元数据连接	根据实际需求，选择 已创建的元数据连接 类型。
数据库名	用户自定义。
数据表名	用户自定义。
源表路径（可选） 元数据连接为Delta 时，为必填项。	输入源端数据表在数据库中的存放路径。
目标表路径（可选） 元数据连接为Delta （无元数据）时，为 必填项。	输入目的端数据表在数据库中的存放路径。

参数	说明
分区信息（可选）	用户自定义。
是否分区表（可选）	根据实际需求，选择该表是否为分区表。
业务负责人(可选)	可以指定数据表所属业务的责任人。
分析责任人（可选）	可以指定数据表的分析责任人。
跨源表	根据实际需求设置。设为跨源表后不再显示该表一致性校验结果。
说明（可选）	您可以输入用于数据表的说明。

步骤6 参数配置完成后，单击“**确认**”按钮，完成数据表新增。在库表清单列表中可以查看新增的数据表。

----结束

批量导入数据表

步骤1 登录[迁移中心](#)管理控制台。

步骤2 在左侧导航栏选择“**调研>大数据血缘**”，进入**大数据血缘**页面。在页面左上角的**当前项目**下拉列表中选择**迁移项目**。

步骤3 在**元数据采集**区域，单击“**查看库表清单**”，进入**元数据采集**页面的**库表清单**页签。

步骤4 单击“**导入数据表**”按钮，弹出**导入数据表**窗口。

步骤5 根据实际需求，选择**已创建的元数据连接**类型。

步骤6 单击“**模板下载**”，将导入模板下载到本地。打开导入模板，将需要添加的数据表对应参数填写到模板内。

须知

- 每个模板最多支持导入10,000张表。
- 待导入的表必须属于同一个元数据来源。
- 元数据连接为**Delta（有元数据）**连接时，模板中的**source_path**参数为必填项。
- 元数据连接为**Delta（无元数据）**连接时，模板中的**source_path**和**target_path**参数为必填项。

步骤7 模板参数填写完成并保存后，单击导入窗口的“**添加文件**”按钮，将填写后的模板文件上传。

步骤8 上传成功后，单击“**确认**”按钮，系统开始导入，并且会在**任务实例管理**页签中的任务实例列表中，新增一条导入数据表的任务实例记录。您可以通过任务实例查看导入状态。

----结束

2.6.3 血缘采集

2.6.3.1 血缘采集概述

通过血缘采集，可以建立起数据元素之间的关系和依赖关系，从而形成一个清晰的数据血缘图。

目前支持血缘采集的文件类型以及采集须知，参见下表。

文件类型	采集须知
DataArts作业	需要提前 导出DataArts作业 到华为云OBS桶，并将桶的读写权限委托给迁移中心。
SQL作业	需要提前导出SQL作业，并上传到华为云OBS桶，再将桶的读写权限委托给迁移中心。
血缘模板	需要下载血缘模板，并填写相应数据。

2.6.3.2 导出作业

本节介绍导出华为云DataArts作业和阿里云DataWorks作业的方法。

导出华为云 DataArts 作业

步骤1 登录DataArts Studio控制台。

步骤2 选择实例，单击“**进入控制台**”，选择对应工作空间。

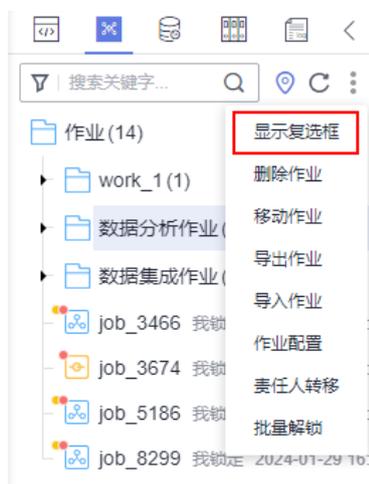
[已开通实例](#) [如何进行准备工作?](#) [快速入门](#) [免费赋能课](#)



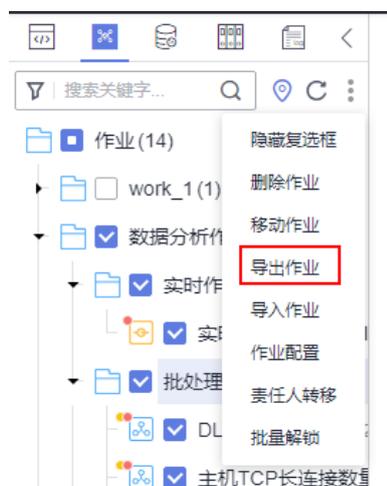
步骤3 单击“**数据开发**”，进入数据开发页面。



步骤4 单击作业目录中的 ，选择“显示复选框”。



步骤5 勾选需要导出的作业，单击  > 导出作业，弹出的导出作业窗口。



步骤6 导出方式选择“导出到OBS”，选择作业保存目录。

步骤7 “导出范围”和“导出状态”根据实际需求选择。

须知

若待导出的作业未提交过任何版本，“导出状态”建议选择“**开发态**”，如果选择了“**提交态**”导出时会跳过该作业。

步骤8 单击“**确定**”，导出完成后，前往**步骤6**选择的OBS目录获取导出结果。

----结束

导出阿里云 DataWorks 作业

步骤1 登录阿里云DataWorks控制台，单击左侧导航栏的“**数据建模与开发 > 数据开发**”，在下拉框中选择对应工作空间后单击“**进入数据开发**”。

步骤2 单击左上方的图标，选择**全部产品 > 更多 > 迁移助手**，进入迁移助手首页。

步骤3 参考**创建和查看DataWorks导出任务**，创建全量导出任务。

----结束

2.6.3.3 创建血缘采集任务

前提条件

已完成**元数据采集**。

操作步骤

步骤1 登录**迁移中心**管理控制台。

步骤2 在左侧导航栏选择“**调研>大数据血缘**”，进入**大数据血缘**页面。在页面左上角的**当前项目**下拉列表中选择**迁移项目**。

步骤3 在**血缘采集**区域，单击“**创建任务**”，弹出**创建血缘采集任务**窗口。



步骤4 根据实际作业类型，选择文件类型并配置对应参数。

文件类型	参数	配置说明
血缘模板	文件	<p>将血缘模板下载到本地，并填写模板内的参数。其中，如下参数值为必填字段。</p> <ul style="list-style-type: none"> ● 目标数据库(TargetDataset) ● 目标表(TargetTable) ● 目标连接名称(TargetConnectionName) ● 目标组件类型(TargetComponentType) ● 上游数据库(SourceDataset) ● 上游表(SourceTable) ● 上游连接名称(SourceConnectionName) ● 上游组件类型(SourceComponentType) ● 作业ID(JobId) <p>须知</p> <ul style="list-style-type: none"> ● 模板中目标组件类型和上游组件类型支持的类型包括 HiveSQL、MaxCompute。 ● 模板中单元格不允许存在公式，否则会导致解析失败。 <p>模板参数填写完成并保存后，单击窗口的“添加文件”，将填写后的模板文件上传至迁移中心。</p> <p>注意 文件大小不能超过100 MB。</p>

- **选择血缘模板。**
 - a. 单击“**模板下载**”，将血缘模板下载到本地。
 - b. 打开血缘模板，填写模板内的参数。其中，如下参数值为必填字段。
 - 目标数据库(TargetDataset)
 - 目标表(TargetTable)
 - 目标连接名称(TargetConnectionName)
 - 目标组件类型(TargetComponentType)
 - 上游数据库(SourceDataset)
 - 上游表(SourceTable)
 - 上游连接名称(SourceConnectionName)
 - 上游组件类型(SourceComponentType)
 - 作业ID(JobId)

须知

- 模板中目标组件类型和上游组件类型支持的类型包括HiveSQL、MaxCompute。
 - 模板中单元格不允许存在公式，否则会导致解析失败。
- c. 模板参数填写完成并保存后，单击窗口的“添加文件”，将填写后的模板文件上传至迁移中心。

注意

文件大小不能超过100 MB。

步骤5 单击“确认”，创建血缘采集任务完成。系统自动开始进行血缘采集。

步骤6 单击“查看任务”，进入血缘采集页面，可以在任务列表中查看采集任务状态。



步骤7 当任务状态为“已完成”时，单击页面右上角的“查看血缘”，可以查看血缘关系图。

----结束

2.6.3.4 查看血缘关系

前提条件

已创建血缘采集任务，且状态为“已完成”。

操作步骤

步骤1 登录迁移中心管理控制台。

步骤2 在左侧导航栏选择“调研>大数据血缘”，进入大数据血缘页面。在页面左上角的当前项目下拉列表中选择迁移项目。

步骤3 在血缘采集区域，单击“查看血缘图”，进入血缘图页面。



步骤4 通过数据库名、表名筛选需要查看的血缘图。如果进行了大数据校验，还可以通过校验批次进行筛选。

----结束

2.7 迁移方案设计

2.7.1 绑定目的端

仅支持主机类型资源绑定目的端。如果您已经在华为云创建了目的端主机，可以将采集的源端主机与目的端主机进行绑定，绑定后，迁移 workflows 将使用所绑定的目的端作为迁移的目的端。

绑定目的端的主机，在创建迁移 workflows 前，无需进行[评估推荐](#)。

前提条件

已完成[资源采集](#)。

注意事项

绑定的目的端主机有以下要求：

- 选择绑定的目的端必须关机，如果需要绑定的目的端处于开机状态，请关机后再绑定。
- 迁移 workflows 将使用所绑定的目的端作为迁移的目的端，迁移过程中该目的端主机会被格式化，然后将源端数据迁移到该主机。
- 如果后续主机迁移 workflows 是公网迁移，所绑定目的端主机需要能访问公网。
- 选择绑定的目的端主机所在区域需要与源端主机[关联的应用](#)处于同一区域。

操作步骤

步骤1 登录[迁移中心](#)管理控制台。

步骤2 单击左侧导航栏的“[设计>迁移方案设计](#)”，进入[迁移方案设计](#)页面。在页面左上角的[当前项目](#)下拉列表中选择[迁移项目](#)。

步骤3 单击目的端规格配置框中的“[目的端配置表](#)”，进入[目的端配置表](#)页面。

步骤4 在[资源配置明细](#)的主机页签，单击主机资源绑定目的端列的“[去绑定](#)”，弹出绑定目的端窗口。

步骤5 选择源端主机关联的[应用](#)所在区域以及区域下的项目，在所选项目下，选择要绑定的目的端主机。支持目的端磁盘缩容，即可以根据源端主机磁盘的实际使用量，选择小于源端磁盘容量和少于源端磁盘数量的目的端主机。

须知

- 当前仅支持Linux主机磁盘缩容，需确保缩容后容量大于源端实际使用量。
- 跨可用区迁移场景仅支持扩容，即使配置缩容，系统也会自动创建一个与源端磁盘容量相同的目的端磁盘。

步骤6 单击“[确认](#)”。目的端绑定成功后，该主机资源绑定目的端列的状态会变为“[已绑定](#)”，单击“[详情](#)”可以查看绑定的目的端规格。并且系统会自动判断所绑定的目的

端是否进行了磁盘扩容，在“**磁盘扩容**”列，“是”代表已扩容；“否”代表未扩容。



----结束

解除目的端绑定

步骤1 在**资源配置明细**的主机页签，单击已绑定目的端的主机资源操作列的“**解除目的端绑定**”，弹出**解除已有目的端**窗口。

步骤2 解绑操作无法恢复，确认解除绑定后，单击“**是**”，完成目的端解绑。

----结束

2.7.2 规格评估

规格评估是根据源端资源规格、性能和应用场景以及特定的成本、可用性、性能、安全合规等需求，为您推荐最合适的华为云对应资源规格，通过规格评估确定迁移目的端。

当前支持规格评估的资源类型为：**主机、数据库、对象存储、容器、中间件**。

说明

已**绑定目的端**的主机资源，无需进行规格推荐，可直接创建迁移工作流。

前提条件

已完成**资源采集**，并将采集的资源关联到**应用**。

操作步骤

步骤1 登录**迁移中心**管理控制台。

步骤2 单击左侧导航栏的“**设计>迁移方案设计**”，进入**迁移方案设计**页面。在页面左上角的**当前项目**下拉列表中选择**迁移项目**。

在迁移方案设计页面，可以查看当前项目下的**资源总数**、资源的**目的端规格配置情况**以及当前项目下所创建的**应用列表**。

步骤3 单击**目的端规格配置框**中的“**规格评估**”按钮，进入**评估单应用**页面。

步骤4 在**选择应用**的下拉列表中，选择需要进行资源规格评估的应用。

步骤5 在**选择资源**区域，勾选应用内，需要评估的资源。

步骤6 根据**表2-38**，配置评估策略。

表 2-38 评估策略参数说明

参数	说明
购买区域	选择在华为云购买资源的区域。推荐您选择靠近目标用户的区域，可以减少网络时延，提高访问速度。
推荐策略	<ul style="list-style-type: none"> ● 匹配源端规格 根据源端规格推荐最合适的华为云对应资源规格，适合云上源端规格推荐。 ● 匹配业务场景 以源端业务场景为基础，结合华为云最佳实践进行资源规格推荐。 ● 跨可用区迁移 仅适用于华为云弹性云服务可用区下线迁移，该策略只会为应用中的主机资源推荐规格。选择该策略同时需要选择目的可用区。 详细的推荐逻辑说明参见主机评估策略推荐逻辑说明。
优先级设置	<ul style="list-style-type: none"> ● 性能优先 推荐的目的端规格要求性能最佳。 ● 价格优先 推荐的目的端规格在满足要求的前提下尽量节约成本。
偏好设置	<p>主机资源，可以进行以下偏好设置，推荐目的端时优先推荐满足偏好设置要求的主机。详细的推荐逻辑说明参见主机评估策略推荐逻辑说明。</p> <ul style="list-style-type: none"> ● 虚拟机类型偏好（可选） 选择期望的目的端主机类型，支持多选。 ● 虚拟机规格族偏好（可选） 根据选择的主机类型，罗列出包含的规格族类型，选择期望的目的端主机规格族，支持多选。 须知 勾选“仅显示适合专属主机规格族”，则只能选择在购买区域所选 Region 下专属主机所包含的规格，并且虚拟机类型偏好将无法选择。 ● 系统盘规格偏好（可选） 选择期望的目的端系统盘规格类型。 ● 数据盘规格偏好（可选） 选择期望的目的端数据盘规格类型。 ● 虚拟机规格调整 选择主机性能规格的推荐策略。 <ul style="list-style-type: none"> - 选择“不调整”，会推荐与源端主机性能相同或最接近的 CPU 核数、内存大小。 - 选择“按性能负载调整”，需要先完成源端主机性能采集，获取源端主机性能负载数据，再通过设置数据处理规则，推荐满足条件的 CPU 核数、内存大小。 须知 采集的性能数据越多，推荐的规格越准确，建议主机性能采集时长不少于 7 天。

参数	说明
	容器资源可以设置“集群类型”、“集群版本”、“容器网络模型”等规格配置，推荐满足规格配置的集群资源。

步骤7 配置完成后，单击“**创建评估**”，主机规格评估创建成功。规格评估完成后，可以**查看评估推荐结果**和**主机性能数据**，其中目的端规格配置即为该资源的迁移目的端。

步骤8 （可选）您还可以进行如下操作：

- **更改目的端规格**：根据实际需求，可以调整目的端主机规格和磁盘规格，并支持磁盘扩容。
- **绑定目的端**：根据评估结果，如果您在华为云上已有满足要求的主机，可以选择并绑定该主机作为目的端主机。
- **绑定专属主机**：如果您需要将源端主机迁移到华为云专属主机 DeH上，可以选择绑定已有专属主机。

----结束

查看评估推荐结果

在**迁移方案设计**页面的**应用列表**，单击操作列的“**查看资源配置**”，进入目的端配置表页面。

在**资源配置明细**区域，可以查看应用内各类型资源的评估推荐结果，其中**目的端规格配置**是根据源端规格配置和评估策略推荐的华为云对应规格配置，**成本预估**是华为云推荐规格所需费用。



查看主机性能数据

在**资源配置明细**的主机列表中，可以查看每台主机最后采集日期7天/30天内的**CPU负载**和**内存负载**平均值。单击“**性能分析**”，可以查看所有主机的性能分布统计。



更改目的端规格

步骤1 在**资源配置明细**区域，单击主机资源操作列的“**更改目的端规格**”，右侧弹出**资源详情**窗口。

步骤2 单击主机规格后的“更改规格”，可以修改目的端主机规格和镜像。

规格配置



步骤3 单击磁盘对应的目的端规格处的“更改规格”，可以修改目的端磁盘的存储类型和资源总用量（磁盘容量）。Linux主机的磁盘容量可以根据源端主机磁盘的实际使用量，进行缩容，即小于源端磁盘容量。更改磁盘规格后，系统会自动判断是否进行了磁盘缩容，在主机资源配置明细列表中的“磁盘缩容”列，“是”代表已缩容；“否”代表未缩容。

须知

- 系统盘容量取值范围为：40 G~1024 G
- 数据盘容量取值范围为：10 G~32768 G
- 当前仅支持Linux主机磁盘缩容，需确保缩容后容量大于源端实际使用量。
- 跨可用区迁移场景仅支持扩容，即使配置缩容，系统也会自动创建一个与源端磁盘容量相同的目的端磁盘。

^ 磁盘 (1)				
是否缩容 否				
资源类型	源端规格	目的端规格	包周期价格/月	按需价格/小时
系统盘	SATA 40 GB	普通IO 40 GB 更改规格	¥12.00	¥0.02

---结束

绑定专属主机

绑定专属主机的作用：

- 在购买资源时，系统会自动将主机购买在您绑定的专属主机上。
- 完成规格评估后，直接创建主机迁移工作流时，系统会在创建目的端主机的那一步自动将主机购买在您绑定的专属主机上。

步骤1 在**资源配置明细**区域，选择主机页签，勾选需要绑定同一专属主机的主机。

步骤2 单击列表上方的“**专属主机配置 > 绑定专属**”，弹出绑定专属主机窗口。

图 2-31 绑定专属主机



步骤3 在下拉列表中选择要绑定的目的端专属主机，单击“**确认**”。绑定成功后，主机资源的**专属主机**列状态会变为“**已绑定**”，单击“**详情**”可以查看绑定的专属主机规格。

----结束

2.7.3 购买资源

目的端规格配置完成后，无需切换服务，可以直接在迁移中心批量购买目的端资源。当前仅支持购买主机资源。主机迁移 workflow 支持自动购买主机资源，通过“**购买列表**”批量购买资源适用于部分主机需要重新部署的场景。通过购买列表成功购买的主机，将自动绑定为迁移目的端，可以使用主机 workflow 继续迁移。

前提条件

已对源端资源完成**规格评估**。

操作步骤

步骤1 登录**迁移中心**管理控制台。

步骤2 单击左侧导航栏的“**设计 > 迁移方案设计**”，进入**迁移方案设计**页面。在页面左上角的**当前项目**下拉列表中选择**迁移项目**。

步骤3 单击**目的端规格配置框**中的“**目的端配置表**”按钮，进入**目的端配置表**页面。

步骤4 在**资源配置明细**区域的**主机**页签，单击已完成规格评估主机购买状态列的“**添加至购买列表**”，购买状态变为“**待购买**”。

步骤5 添加完成后，在页面上方的**购买列表框**中，单击“**查看列表**”按钮，进入**购买列表**页面。

步骤6 在**操作流程**区域，单击“**查看我的模板**”，右侧弹出“**我的模板**”窗口，单击“**创建模板**”按钮，进入**创建配置模板**窗口。

步骤7 在**基础信息**区域，选择模板类型（当前仅支持主机），自定义模板名称；在**配置信息**区域，根据**表2-39**，配置模板参数。

表 2-39 模板参数配置说明

参数	配置说明
区域	选择目的端区域。
项目	选择目的端区域项目。
可用区	默认随机分配，也可手动选择。
虚拟私有云	选择目的端区域创建的虚拟私有云（VPC）。
子网	选择虚拟私有云所包含的子网，子网网段必须在所选VPC网段范围内。
安全组	所选安全组需要满足： <ul style="list-style-type: none"> Windows系统开放8899端口、8900端口和22端口。 Linux系统开放22端口。
企业项目	将购买的主机加入商用生产企业项目。默认项目为default。 企业项目管理提供了一种按企业项目管理云资源的方式，帮助您实现以企业项目为基本单元的资源及人员的统一管理。关于创建和管理企业项目的详情，请参见《 企业管理用户指南 》。

- 步骤8** 参数配置完成后，单击“保存”按钮，在模板列表可以查看创建的模板。
- 步骤9** 模板创建完成后，返回购买列表，单击主机购买模板列的“绑定模板”，弹出“绑定购买模板”窗口，在下拉列表中选择创建的模板，单击“确认”。
- 步骤10** 为主机绑定购买模板后，单击操作列的“去购买”，当购买状态为“购买成功”时，可以看到购买资源的ID。

----结束

2.7.4 迁移风险评估

在资源迁移前，可以通过风险评估，提前检查可能影响迁移的风险项，包括但不限于兼容性问题、固件兼容性、规格差异等。

须知

目前仅支持主机资源进行迁移风险评估。

前提条件

已通过公网发现、内网发现或手动添加中的任意一种方法获取主机资源。

操作步骤

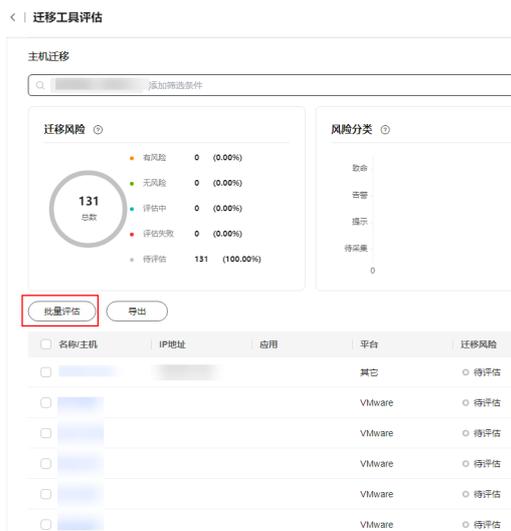
- 步骤1** 登录[迁移中心控制台](#)。
- 步骤2** 单击左侧导航栏的“设计>迁移方案设计”，进入迁移方案设计页面。在页面左上角的当前项目下拉列表中选择[迁移项目](#)。

步骤3 单击迁移工具评估框中的“查看评估表”，进入迁移工具评估页面。



步骤4 在主机页签，可以看到所有主机资源列表。

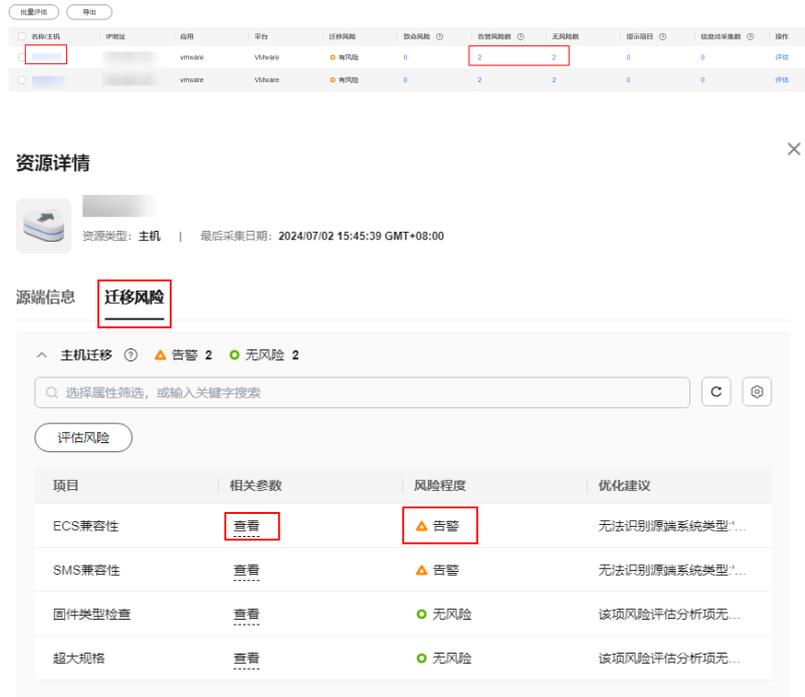
- 如果需要评估列表中所有“待评估”状态的主机，单击“批量评估”按钮，在弹出的对话框中单击“确定”，即可对所有“待评估”状态的主机进行风险评估。



- 如果仅需要评估部分主机，可以在列表中勾选需要评估的主机；也可以通过上方的搜索框，根据主机名称、IP地址、所属平台、关联的应用等条件过滤列表中的主机资源，然后勾选需要评估的主机。勾选完成后，单击“批量评估”按钮，在弹出的对话框中确认待评估主机数量无误后，单击“确认”，即可对勾选的主机进行风险评估。



步骤5 在评估列表中可以查看各主机的迁移风险评估状态，以及不同风险级别存在的风险项数量。单击主机名称或风险项数量，右侧弹出资源详情窗口，在迁移风险页签，可以查看评估包含的所有项目以及各项目的参数说明和风险程度。



步骤6 对于存在风险的项目，可以根据优化建议进行处理，处理完成后，在评估列表中，单击操作列的“评估”，可再次对该主机进行风险评估。



----结束

2.8 迁移方案配置

2.8.1 配置主机迁移方案

主机迁移方案是面向主机迁移 workflow 进行的一站式配置，能够实现批量评估和配置目的端主机规格。本节介绍为主机迁移配置迁移方案的方法。

准备工作

- 创建[迁移项目](#)。
- 通过资源采集，完成主机资源的发现与采集。MgC提供了三种主机采集方式，您可以根据实际场景和您源端主机所在的云平台，选择适合的采集方式。
 - 源端主机位于云平台：源端为阿里云、华为云、AWS、腾讯云、谷歌云、Azure等云平台主机，推荐使用[公网采集](#)，也可以使用各场景通用的[手动添加](#)方式。
 - 源端主机位于线下IDC：推荐使用[内网采集](#)，通过[网段扫描](#)或[VMware扫描](#)，采集源端主机资源，也可以使用各场景通用的[手动添加](#)方式。
- 在源端内网环境中安装迁移工具Edge并注册用户，方法和要求请参见[安装Windows版本的Edge](#)。安装并注册成功后，将Edge与云端MgC控制台建立连接，方法请参考[连接迁移中心](#)。

约束与限制

- 一个方案只能创建一个迁移工作流。
- 同一台主机只能添加到一个方案内。
- 一个方案最多添加 100 台主机。

操作步骤

步骤1 登录**迁移中心**管理控制台。在页面左上角的**当前项目**下拉列表中选择在准备工作中创建的迁移项目。

步骤2 在左侧导航栏选择“**设计>迁移方案配置**”，进入**迁移方案配置**页面。

步骤3 单击页面右上角的“**创建迁移方案**”按钮，右侧弹出**迁移方案模板**窗口。

步骤4 单击**主机标准迁移**框中的“**开始配置**”按钮，进入**主机标准迁移**页面。



步骤5 在**基础配置**区域，根据**表2-40**，配置相关参数。

表 2-40 基础配置参数配置说明

参数	配置说明
方案名称	用户自定义。
描述（可选）	
源端平台	选择源端主机所在的平台。 须知 同一方案下的主机需要来自同一平台。
目的端区域	选择需要迁移到的目的端区域。

步骤6 在**待迁移主机列表**区域，单击“**添加主机**”按钮，右侧弹出**添加主机**窗口。



步骤7 勾选需要迁移的主机，单击“确认”。

注意

- 同一方案下的待迁移主机，将通过同一工作流迁移，届时将使用同样迁移设置，如 VPC、子网、安全组等，请自行确认该方案下主机资源范围。
- 所选的主机必须都来源于**步骤4**.基础配置所选的源端平台。

步骤8 将待迁移的主机添加完后，单击“保存并下一步”，进入**准备主机**阶段。

步骤9 添加主机凭证。

- **手动添加**：适用于主机数量较少的场景。添加方法请参考**添加资源凭证**。
- **批量添加**：适用于主机数量较多的场景。
 - a. 单击待迁移主机列表上方的“**批量绑定凭证**”按钮，下载批量绑定凭证模板文件。打开模板文件，填写所有待迁移主机的凭证信息（用户名/密码）后保存。



- b. 将保存后的凭证模板文件导入与**MgC控制台连接的Edge**上，系统将自动为主机绑定迁移代理和凭证。



为主机添加凭证后，您还可以单击操作列的“**修改**”，调整接入IP、接入端口、迁移代理和凭证等参数值。

步骤10 在待迁移主机列表中，勾选所有主机，单击列表上方的“**批量准备度检查**”按钮，系统会对待迁移的主机进行迁移准备度检查。



- 当**准备度检查**列状态会变为“**已完成**”时，代表该主机已满足迁移条件。

- 当**准备度检查**列状态提示“缺少必要信息”时，请单击深度采集列的“采集”，进行深度采集。
- 当**准备度检查**列状态提示**网络相关问题**时，请检查您填写的接入参数是否正确，确认无误后重新进行准备度检查。

步骤11 当所有待迁移主机的准备度检查状态均为“已完成”后，单击“下一步”，进入可选步骤“性能采集”。

步骤12 （可选）主机性能采集。通过采集源端主机的实际性能负载，在后续的规格推荐中，自动调整目的端主机规格。**性能采集通过在源端抽样获取，每5分钟抽样一次，需要较长时间，请自行选择是否执行。**更多约束与限制请查看**性能采集使用须知**。如果不需要进行主机性能采集，直接单击“下一步”，进入“目的端配置”阶段。

勾选需要进行性能采集的主机，单击列表上方的“批量性能采集 > 开始”，启动主机性能采集。此时性能采集列的状态为“采集中”。

开始性能采集后，您还可以进行如下操作：

- 单击操作列的“查看性能详情”，可以查看采集到的主机性能数据。
- 单击列表上方的“批量性能采集 > 停止”，可以批量暂停性能数据采集。

步骤13 配置目的端。您可以根据实际情况，选择配置目的端主机的方法。

- **绑定目的端。**如果您已在华为云购买了目的端主机，可以将源端主机与目的端主机进行绑定，后续创建的迁移 workflow 将使用所绑定的主机作为迁移目的端。关于绑定的目的端主机要求请参见**绑定目的端注意事项**。

a. 在待迁移主机列表，单击绑定目的端列的“去绑定”，弹出**绑定目的端**窗口。



b. 选择目的端主机所在区域以及区域下的项目，在所选项目下，选择要绑定的目的端主机。支持目的端磁盘扩容，即可以根据源端主机磁盘的实际使用量，选择小于源端磁盘容量和少于源端磁盘数量的目的端主机。

须知

- 请确保目的端主机绑定了弹性公网IP。
 - 当前仅支持Linux主机磁盘扩容，需确保扩容后容量大于源端实际使用量。
 - 跨可用区迁移场景仅支持扩容，即使配置扩容，系统也会自动创建一个与源端磁盘容量相同的**目的端磁盘**。
- c. 单击“确认”。目的端绑定成功后，该主机资源**绑定目的端**列的状态会变为“已绑定”，单击“详情”可以查看绑定的目的端规格。并且系统会自动判断所绑定的目的端是否进行了磁盘扩容，在“**磁盘扩容**”列，“是”代表已扩容；“否”代表未扩容。
- **批量规格推荐。**根据源端主机规格，系统会自动推荐在华为云与之对应的主机规格作为目的端，您还可以对推荐的主机规格进行手动调整。后续创建的 workflow 会

按所配置规格自动创建目的端主机。已绑定目的端的主机资源，无需进行规格推荐。

- a. 在待迁移主机列表，勾选需要进行规格推荐的主机，单击列表上方的“批量规格推荐”，右侧弹出批量规格评估窗口。
- b. 根据表2-41，配置评估策略。

表 2-41 评估策略参数说明

参数	说明
优先级设置	<ul style="list-style-type: none"> ▪ 性能优先 推荐的目的端规格要求性能最佳。 ▪ 价格优先 推荐的目的端规格在满足要求的前提下尽量节约成本。
偏好设置	<p>推荐目的端时优先推荐满足偏好设置要求的主机。详细的推荐逻辑说明参见主机评估策略推荐逻辑说明。</p> <ul style="list-style-type: none"> ▪ 虚拟机类型偏好（可选） 选择期望的目的端主机类型，支持多选。 ▪ 虚拟机规格族偏好（可选） 根据选择的主机类型，罗列出包含的规格族类型，选择期望的目的端主机规格族，支持多选。 <p>须知 勾选“仅显示适合专属主机规格族”，则只能选择在购买区域所选Region下专属主机所包含的规格，并且虚拟机类型偏好将无法选择。</p> <ul style="list-style-type: none"> ▪ 系统盘规格偏好（可选） 选择期望的目的端系统盘规格类型。 ▪ 数据盘规格偏好（可选） 选择期望的目的端数据盘规格类型。 ▪ 虚拟机规格调整 选择主机性能规格的推荐策略。 <ul style="list-style-type: none"> ○ 选择“不调整”，会推荐与源端主机性能相同或最接近的CPU核数、内存大小。 ○ 选择“按性能负载调整”，需要先完成源端主机性能采集，获取源端主机性能负载数据，再通过设置数据处理规则，推荐满足条件的CPU核数、内存大小。 <p>须知 采集的性能数据越多，推荐的规格越准确，建议主机性能采集时长不少于7天。</p>

- c. 配置完成后，单击“确认”，系统开始进行主机规格评估。当目的端配置状态为“成功”时，代表规格评估完成，目的端规格即为该主机目的端规格。



- d. (可选) 更改目的端规格。您可以对推荐的目的端主机规格进行手动调整。
- i. 单击主机资源操作列的“更改目的端规格”，右侧弹出资源详情窗口。
 - ii. 单击主机规格后的“更改规格”，可以修改目的端主机规格和镜像。

规格配置



- iii. 单击磁盘对应的目的端规格处的“更改规格”，可以修改目的端磁盘的存储类型和资源总用量（磁盘容量）。Linux主机的磁盘容量可以根据源端主机磁盘的实际使用量，进行缩容，即小于源端磁盘容量。更改磁盘规格后，系统会自动判断是否进行了磁盘缩容，在主机资源配置明细列表中的“磁盘缩容”列，“是”代表已缩容；“否”代表未缩容。

须知

- 系统盘容量取值范围为：40 G~1024 G
- 数据盘容量取值范围为：10 G~32768 G
- 当前仅支持Linux主机磁盘缩容，需确保缩容后容量大于源端实际使用量。
- 跨可用区迁移场景仅支持扩容，即使配置缩容，系统也会自动创建一个与源端磁盘容量相同的目的端磁盘。

磁盘 (1)				
是否缩容 否				
资源类型	源端规格	目的端规格	包周期价格/月	按需价格/小时
系统盘	SATA 40 GB	普通IO 40 GB 更改规格	¥12.00	¥0.02

- **批量配置规格。**可以批量为待迁移主机设置同一规格的目的端主机。

注意

批量配置规格为手动配置，操作之前需要您自行评估并承担相关风险。请确保您选择的规格完全符合您的系统需求和兼容性标准。

- a. 在待迁移主机列表，勾选需要迁移到同一规格的主机，单击列表上方的“**批量配置规格**”，右侧弹出**批量配置规格**窗口。
- b. 选择目的端主机规格后，单击“**确认**”。设置成功后主机的“目的端配置状态”会显示为“**手动设置**”。



- **批量设置镜像**。可以批量为待迁移主机设置同一类型的镜像。

注意

批量设置镜像为手动配置，需要您自行评估并承担相关风险。请确保您选择的镜像完全符合您的系统需求和兼容性标准。

- a. 在待迁移主机列表，勾选需要设置为同一种镜像类型的主机，单击列表上方的“**批量设置镜像**”，弹出**批量设置镜像**窗口。
- b. 根据实际需求，选择镜像类型和规格，单击“**确认**”。请确保目的端主机的系统盘容量大于镜像所需空间。

- **高级批量配置**。通过导出、导入的方式，批量进行目的端规格、磁盘、镜像设置，适合专业迁移服务人员。

- a. 在待迁移主机列表，单击列表上方“**高级批量配置 > 导出目的端配置**”，将列表中的所有主机信息以Excel文件格式下载到本地。



- b. 打开导出的文件，选择第一个sheet页，名为“**servers**”。其中蓝色列（A~G列）为源端主机信息，**无需修改**；黄色列（H~K列）为目的端主机信息，**需要配置**。

Host Name	Source Specification	Destination Specification	Batch Count	Batch Name	Batch Status	Batch Type
192.168.1.1	OS: Linux 2.2 4 GB	OS: Linux 2.2 4 GB	1	batch1	手动设置	镜像
192.168.1.2	OS: Linux 2.2 4 GB	OS: Linux 2.2 4 GB	1	batch2	手动设置	镜像

- c. 根据**表2-42**，配置目的端主机信息（H~K列）。

表 2-42 H~K 列参数配置说明

参数	配置说明
flavor	填入目的端主机规格名称。在第二个sheet页（名为flavors），列出了目的端区域在售的 S（通用型）和C（计算增强型）系列的规格。选择合适的规格，将规格名称复制到该参数单元格。 说明 如果进行过规格评估，导出结果包含评估结果；如果未进行过规格评估，导出结果为空。

参数	配置说明
disks(name,osDisk,type,capacity;)	填入所有磁盘信息，多个磁盘之间以英文分号 (;) 隔开。每块磁盘必须包含 名称 ， 是否为系统盘 ， 磁盘类型 ， 磁盘大小 等信息，信息之间以英文逗号 (,) 隔开。磁盘大小只写数字，单位默认是GiB。 例如： /dev/sda,true,SAS,1024;/dev/sdb,false,SAS,2046 说明 如果进行过规格评估，导出结果包含评估结果；如果未进行过规格评估，磁盘命名为源端磁盘名称，磁盘类型为SAS，磁盘大小为源端磁盘大小。
image_type	填入镜像类型。 <ul style="list-style-type: none"> public (公共镜像) private (私人镜像) shared (共享镜像)
image	填入镜像名称。可以在第三个sheet页 (名为 images)，查看目的端区域所有可用的镜像。

- d. 所有主机的目的端信息配置完成后，保存Excel文件。返回MgC控制台，单击待迁移主机列表上方的“高级批量配置 > 导入目的端配置”，弹出导入目的端配置窗口。



- e. 单击窗口中的“前往OBS存放文件”，跳转到对象存储服务 OBS的桶列表页面。



- f. 选择要存放目的端配置文件 (Excel文件) 的桶，将Excel文件上传到该桶中。
g. 上传成功后，返回MgC控制台，在导入目的端配置窗口的下拉列表中，选择存放在OBS桶中的目的端配置文件。

- h. 单击“确认”导入目的端配置文件。

📖 说明

- flavor参数为空时，忽略该行。
- disk或image参数为空时，保持原有配置。
- disk数量跟源端不一致时，会报磁盘数量不一致错误。
- 磁盘类型跟flavor冲突时，会报磁盘类型错误。
- **专属主机配置。**如果您需要将源端主机迁移到华为云专属主机 DeH上，可以选择绑定已有专属主机。绑定后，主机迁移工作流会在创建目的端主机的那一步自动将主机购买在您绑定的专属主机上。
 - a. 在待迁移主机列表，勾选需要绑定专属主机的主机。
 - b. 单击列表上方的“**专属主机配置 > 绑定专属**”，弹出绑定专属主机窗口。
 - c. 在下拉列表中选择要绑定的目的端专属主机，单击“**确认**”。绑定成功后，该主机**专属主机**列的状态会变为“**已绑定**”，单击“**详情**”可以查看绑定的专属主机规格。

步骤14 所有待迁移主机的目的端配置完成后，根据需求选择后续操作。

- **立即创建主机迁移工作流**
单击“**创建工作流**”按钮，进行迁移工作流参数配置。参数配置说明请参见[主机迁移工作流参数配置说明](#)。
- **仅保存迁移方案**
单击“**确定**”，将该方案保存在迁移方案列表中。可以对保存的迁移方案进行如下操作：
 - **修改迁移方案：**单击操作列的“**设计**”，可以对方案中的配置进行修改。
 - **创建主机迁移工作流：**单击操作列的“**创建工作流**”，使用该方案创建主机迁移工作流。参数配置说明请参见[主机迁移工作流参数配置说明](#)。

---结束

2.8.2 配置存储批量迁移方案

迁移中心为多种资源的不同迁移方案配置模板，可以面向迁移工作流进行一站式配置，并按应用显示配置进度。本节介绍为对象存储批量迁移配置迁移方案的方法。

对象存储批量迁移：适用于源端有多个对象存储桶资源需要迁移的场景。如果只有单个对象存储桶资源需要迁移，请直接[创建单个对象存储迁移工作流](#)。

前提条件

- 已通过资源采集的公网采集，完成对象存储资源的发现与采集。
- （可选）已对待迁移的对象存储资源进行深度采集。

注意事项

- 创建方案时，待迁移桶列表中的资源所属的平台必须和所选的源端平台一致，且一个方案对应一个目的端区域。
- 一个源端桶资源可以关联多个方案。

- 在一个方案下，一个桶资源的目的端前缀只能有一个，但是源端前缀可以有多个。

操作步骤

- 步骤1** 登录[迁移中心](#)管理控制台。在左侧导航栏选择“设计>迁移方案配置”，进入迁移方案配置页面。
- 步骤2** 单击页面右上角的“创建迁移方案”按钮，右侧弹出迁移方案模板窗口。
- 步骤3** 单击对象存储批量迁移框中的“开始配置”按钮，进入对象存储批量迁移页面。
- 步骤4** 在基础配置区域，根据[表2-43](#)，配置相关参数。

表 2-43 基础配置参数配置说明

参数	配置说明
方案名称	用户自定义。
描述（可选）	用户自定义。
源端平台	选择源端资源所在的云平台。
目的端桶区域	选择目的端桶所在区域。

- 步骤5** 在待迁移桶列表区域，单击“添加桶”按钮，右侧弹出添加桶窗口。
- 步骤6** 勾选需要迁移的桶，单击操作列的“修改”，调整桶的迁移方式，迁移方式说明参见[表2-44](#)。勾选并调整完成后，单击“确认”。

注意

- 所选的资源必须都来源于基础配置所选的源端平台。
- 最多添加 100 个桶。

表 2-44 迁移方式说明

迁移方式	说明
全量迁移	迁移源端桶/文件夹内的全部数据。
列表迁移	通过列表文件过滤迁移，只迁移列表文件中的对象。列表文件需要存放在目的端桶中。
前缀迁移	通过对象名称或者名称前缀过滤迁移，只迁移与对象名称或名称前缀一致的对象。

- 步骤7** 在待迁移桶列表中，单击操作列的“修改”，弹出修改迁移设置窗口。
- 步骤8** 选择源端凭证，输入对象个数和容量（可选），单击“确认”，完成桶的迁移设置。

- 如果桶的迁移方式为“**列表迁移**”，需要在**列表路径**中输入列表文件的存放路径并以“/”结尾。
- 如果桶的迁移方式为“**前缀迁移**”，在“**前缀路径**”添加对象名称或者名称前缀，单击“**确定**”保存前缀。

说明

输入对象前缀路径时，如果对象文件在源端桶的根目录下，则直接添加前缀；如果对象文件存在目录路径，则需要对象前缀前添加对象文件所在的目录路径。格式为：文件夹/对象前缀

- 步骤9** 所有待迁移桶的迁移设置完成后，单击“**下一步**”，进行目的端配置。
- 步骤10** 单击操作列的“**修改**”，选择目的端凭证和目的端桶，输入目的端前缀（可选），单击“**保存**”。
- 步骤11** 所有资源的目的端配置完成后，单击“**下一步**”，进行集群评估。集群评估为可选操作，您可以跳过该步骤，直接绑定已有集群，也可以通过评估，获取集群规格推荐后，去创建集群。集群的创建方法请参考[创建集群](#)。
- 步骤12** 单击“**下一步**”，进入集群配置，绑定已有迁移集群，本方案下所有资源将通过此集群迁移。
- 步骤13** 单击“**立即绑定**”，右侧弹出**集群选择**窗口。
- 步骤14** 在集群列表中，选择需要使用的集群，单击“**确认**”。如果需要对集群进行修改，请参考[集群管理](#)。
- 步骤15** 集群配置完成后，单击“**确定**”，迁移方案创建完成。在**迁移方案配置**页面的**方案列表**，可以查看已创建的迁移方案。
- 如果需要对迁移方案配置进行调整，可以单击操作列的“**设计**”进行修改。
 - 当**设计进度**为已完成时，单击操作列的“**创建工作流**”，即可[创建对象存储批量迁移工作流](#)。

---结束

2.8.3 配置文件存储批量迁移方案

本节介绍为文件存储批量迁移配置迁移方案的方法。

文件存储批量迁移：适用于源端有多个文件存储地址内的资源需要迁移的场景。如果只有单个文件存储地址内的资源需要迁移，请直接[创建单个文件存储迁移工作流](#)。

准备工作

- **创建迁移项目。**
在MgC控制台为本次迁移创建独立的项目，方法请参见[项目管理](#)。
- **创建迁移集群。**
通过集群可以创建Master节点、迁移节点和列举节点，确保文件存储工作流的顺利运行。创建方法请参考[创建集群](#)。
- **创建目的端弹性文件系统。**
在目的端创建相应的弹性文件系统，用于迁移后的数据存储和管理，方法请参见[创建弹性文件系统](#)或[创建高性能文件系统](#)。

注意事项

- 一个方案只能创建一个迁移工作流。
- 一个方案最多添加100个文件存储。
- 同一方案下的目的端文件存储必须处于同一区域。

操作步骤

- 步骤1** 登录**迁移中心管理控制台**。
- 步骤2** 在左侧导航栏选择“设计>迁移方案配置”，进入**迁移方案配置**页面。
- 步骤3** 单击页面右上角的“创建迁移方案”按钮，右侧弹出**迁移方案模板**窗口。
- 步骤4** 单击**文件存储批量迁移**框中的“开始配置”按钮，进入**文件存储批量迁移**页面。
- 步骤5** 在**基础信息**区域，根据**表2-45**，配置相关参数。

表 2-45 基础信息参数配置说明

参数	配置说明
方案名称	用户自定义。
描述（可选）	用户自定义。
目的端文件存储区域	选择目的端文件存储所在区域。

- 步骤6** 单击“添加文件存储”按钮，在下方的列表中会新增一行源端文件存储配置信息。

图 2-32 添加源端文件存储



- 步骤7** 根据**表2-46**，配置源端文件存储信息。

表 2-46 源端文件存储参数配置说明

参数	说明	须知
存储类型	支持的源端存储类型包括： NAS_GFS、 NAS_NFS_V3_PROTOCOL、 NAS_NFS_V3_MOUNT，请根据 实际情况选择。	-

参数	说明	须知
迁移方式	<p>选择迁移方式。</p> <ul style="list-style-type: none"> 全量迁移：迁移源端文件存储服务地址+路径内的全部数据。 列表迁移：通过列表文件过滤迁移，只迁移列表文件中的对象。 	-
列表路径	<p>迁移方式选择列表迁移时，填写列表文件所在路径。迁移方式选择全量迁移时，无需填写。</p>	<p>列表文件需要满足如下要求：</p> <ul style="list-style-type: none"> 列表文件大小不能超过30 MB。 列表文件类型必须为.txt类型。 列表文件必须是UTF-8无BOM格式编码格式。 列表文件每行只能包含一个文件名称，且文件名称使用URL Encode编码。 列表文件行长度不超过$16 \times 1024 = 16384$字节，否则迁移失败。 列表文件行数不超过10000
文件存储服务地址	<p>输入源端文件系统的挂载地址。</p> <ul style="list-style-type: none"> 存储类型为NAS_GFS时，文件存储服务地址可以填写主机名、域名、可选的子域、灵活的端口号和路径表示。 存储类型为NAS_NFS_V3_PROTOCOL和NAS_NFS_V3_MOUNT时，格式为： <ul style="list-style-type: none"> IP地址:/ IP地址:/ xxx 域名:/ 域名:/ xxx 	<p>无论是全量迁移还是列表迁移，源端文件存储服务地址+路径的组合或源端文件存储服务地址+列表路径的组合都必须是唯一的不能重复，都遵循前缀匹配规则。</p> <p>例如，已有组合是地址192.168.0.1:/a和路径/abc/a，那么添加或修改另一个地址为192.168.0.1:/a和路径/abc时，会因重复而报错。要覆盖已有路径，需要先删除旧的地址和路径组合。</p>
路径	<p>输入待迁移文件路径，格式为：/文件夹。</p>	<p>列表迁移时，路径填写为/即可。</p>

步骤8 配置完成后，单击操作列的“保存”。所有待迁移文件存储添加完成后，单击“下一步”，进行目的端配置。

步骤9 在列表中单击操作列的“修改”，根据表2-47，配置目的端文件存储信息。

表 2-47 目的端文件存储参数配置说明

参数	说明	须知
源端路径	根据源端文件存储服务地址+路径的组合自动生成。	-
存储类型	目前只支持 NAS_NFS_V3_MOUNT。	-
文件存储服务地址	<ul style="list-style-type: none"> 产品类型选择通用文件系统时，输入目的端文件系统的挂载地址。挂载地址可以在文件系统列表页面，单击挂载地址后面的  图标复制获取。 产品类型选择SFS Turbo时，输入目的端文件系统的共享路径。共享路径可以在SFS Turbo文件系统列表页面，单击共享路径后面的  图标复制获取。 	文件存储服务地址+路径的组合可以重复。
路径	输入目的端接收文件的路径，格式为：/文件夹。	-

步骤10 配置完成后，单击操作列的“保存”。所有资源的目的端配置完成后，单击“下一步”，进行集群评估。集群评估为可选操作，您可以跳过该步骤，直接绑定已有集群，也可以通过评估，获取集群规格推荐后，去创建集群。集群的创建方法请参考[创建集群](#)。

步骤11 单击“下一步”，进入集群配置，绑定已有迁移集群，本方案下所有资源将通过绑定的集群迁移。

步骤12 单击“立即绑定”，右侧弹出**集群选择**窗口。

步骤13 在集群列表中，选择需要使用的集群（要求集群状态为健康），单击“确定”。如果需要~~对~~对集群进行修改，请参考[集群管理](#)。

图 2-33 选择集群



步骤14 集群配置完成后，单击“确定”，迁移方案创建完成。在**迁移方案配置**页面的**方案列表**，可以查看已创建的迁移方案。

- 如果需要对迁移方案配置进行调整，可以单击操作列的“设计”进行修改。
- 当设计进度为已完成时，单击操作列的“创建工作流”，即可[创建文件存储批量迁移工作流](#)。

----结束

2.9 迁移集群

2.9.1 创建集群

迁移集群是专为存储工作流提供协作中的核心组件，通过集群可以创建和管理迁移节点、列举节点，部署和升级迁移插件，确保存储工作流顺利运行。

使用须知

- 单个账号在每个区域（Region）最多只能创建2个集群。
- 为了保证迁移的稳定性和数据安全，创建的迁移集群不支持用户自行登录。如您有特殊需求必须登录，请联系技术支持人员。
- 当集群超过30天未使用时会被回收（不会删除您的资源，但是集群将变为“已废弃”状态，且不支持恢复）。
- 迁移集群所包含的资源类型以及所有节点的安全组配置详细信息请查看[资源类型和安全组配置说明](#)。

费用说明

在创建迁移集群时，会产生以下费用：

- 创建Master节点、迁移节点和列举节点会产生ECS服务费用。具体计费详情请参见[ECS按需计费说明](#)或[ECS价格计算器](#)。
- 如果您使用公网迁移时，会产生NAT网关服务费用。计费详情请参见[NAT网关计费说明](#)或[NAT价格计算器](#)。
- 启用LTS服务时，会产生LTS服务费用。具体计费详情请参见[LTS计费说明](#)或[LTS价格计算器](#)。

操作步骤

步骤1 登录[迁移中心管理控制台](#)。

步骤2 在左侧导航栏选择“部署>迁移集群”，进入迁移集群页面。

步骤3 单击页面右上角的“创建集群”按钮，进入创建集群页面。

如果是首次创建迁移集群，需要您同意华为云账号进行相关的委托授权。委托权限所包含的细粒度权限请前往[委托权限说明](#)查看。



步骤4 根据表2-48，配置参数。

表 2-48 集群参数配置说明

区域	参数	配置说明	限制说明
基础配置	集群名称	自定义名称。	同一账户下集群名称不可重名。
	区域	选择创建集群的区域。	与目的端区域保持一致。
	集群类型	选择集群使用场景。	当前仅支持存储迁移场景。
节点创建	Master节点ECS规格	用于管理迁移节点和列举节点。	节点数量为1，无法修改。
	迁移节点ECS规格	迁移节点用于执行迁移和对比的动作，推荐规格8U16G。	<ul style="list-style-type: none"> 规格选定后，后续无法更改。 设置的节点数量需要同时满足如下要求： <ul style="list-style-type: none"> 迁移节点数+列举节点数+1 ≤ 100 迁移节点数+列举节点数+1 ≤ 子网可用IP数
	列举节点ECS规格	列举节点用于执行列举动作，推荐规格8U16G。	
网络配置	虚拟私有云	在下拉列表中选择虚拟私有云。	-
	子网	选择与虚拟私有云网段相同且满足可用IP数的子网。	子网的可用IP数≥迁移节点数+列举节点数+1。

区域	参数	配置说明	限制说明
	模式选择	<ul style="list-style-type: none"> 公网：使用公网迁移，并选择公网网关。如果没有可用网关，选择“新建网关”，选择网关规格和EIP（EIP最多支持同时勾选20个）。 内网：使用内网迁移，适用于同一区域内的迁移场景。 专线：使用专线迁移，通过专线直接访问源端数据。搭建云专线的方法可参考云专线 DC。 	-
高级设置	DNS配置（可选）	填写DNS服务器IP地址，用于追加/etc/resolv.conf文件中的nameserver参数值。多个DNS服务器地址使用英文逗号分隔，例如：192.0.2.1,192.0.2.2	最多可输入3个DNS服务器地址。
	域名映射配置（可选）	添加域名到IP地址的映射，用于修改/etc/hosts文件。	最多可添加500个IP地址映射。
-	流量控制	<p>设置迁移任务在指定时段的最大流量带宽。</p> <ul style="list-style-type: none"> 不勾选，则不进行流量控制 勾选，进行流量控制。在“开始时间”、“结束时间”和“限速”文本框中输入数值。 <p>须知 例如：设置开始时间08:00、结束时间12:00、限制最大流量20 MB/s，当迁移任务运行到（08:00~12:00）该时间段时，最大迁移速度将限制为20 MB/s。该时间段以外，迁移速度不做限制。</p>	<ul style="list-style-type: none"> 最多可添加五条流量控制规则。 时区时间以控制台所选Region的当地时间为准。
	启用LTS服务	<ul style="list-style-type: none"> 启用：记录存储迁移过程中的日志，便于查看和定位迁移中出现的问题。 不启用：不记录存储迁移过程中的日志。 	-

步骤5 参数配置完成后，单击“**确认**”开始创建集群。在迁移集群列表中查看集群状态，状态详情参见**集群状态说明**。如果集群状态为创建失败，将鼠标移动到“创建失败”上，可以查看失败原因。在解决失败原因后，可以通过操作列的“更多 > 重试”功能，重新创建集群。

步骤6 迁移集群创建成功后，您可以进行如下操作。

- 当集群状态为**健康**或者**亚健康**时，单击集群操作列的，可以配置存储迁移 workflow。
- 在**集群管理**界面中，您可以添加/删除节点、流量控制规则等。

----结束

升级插件版本

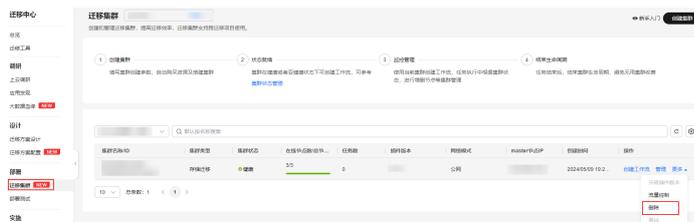
当集群的插件版本有更新时，可以一键升级集群插件版本。在集群列表中，单击操作列的“更多 > 升级插件版本”，即可将集群插件升级为最新版本。



删除集群

当您完成数据迁移并且集群不再需要时，可以直接删除闲置的集群以释放资源。

步骤1 在集群列表，选择需要清理的集群，单击操作列的“更多 > 删除”，弹出**删除集群**窗口。



步骤2 选择是否同步删除关联资源后，单击“**确认**”，删除集群。



----结束

重试

对于创建失败的集群，在解决失败原因后，可以尝试重新创建集群。在集群列表中，单击操作列的“更多 > 重试”，即可重新创建集群资源。



2.9.2 集群管理

在集群创建完成后，您可以根据实际需求随时调整集群内的流量控制规则和节点信息。

前提条件

已完成[集群创建](#)。

添加流量控制规则

- 步骤1** 登录[迁移中心](#)管理控制台。在左侧导航栏选择“部署>迁移集群”，进入[迁移集群](#)页面。
- 步骤2** 在集群列表中，单击操作列的“管理”，右侧弹出集群详细信息页面。
- 步骤3** 在[流量控制](#)区域，单击“添加”，下方新增一条规则记录。
- 步骤4** 在“开始时间”、“结束时间”和“限速”文本框中输入数值。

例如：设置开始时间08:00、结束时间12:00、限制最大流量20 MB/s，当迁移任务运行到（08:00~12:00）该时间段时，最大迁移速度将限制为20 MB/s。该时间段以外，迁移速度不做限制。

须知

- 时区时间以控制台所选Region的当地时间为准。
- 最多可以创建5条记录。

步骤5 单击“确认”，完成流量控制规则的添加。

----结束

新增节点

新增迁移节点和列举节点的操作步骤相同，以下步骤以新增迁移节点为例进行介绍。

步骤1 登录**迁移中心**管理控制台。在左侧导航栏选择“部署>迁移集群”，进入**迁移集群**页面。

步骤2 在集群列表中，单击操作列的“管理”，右侧弹出集群详细信息页面。



步骤3 在节点信息区域，单击“新增节点>迁移节点”，弹出添加节点页面。



步骤4 根据实际需求，选择节点的ECS规格，并输入需要增加的节点数量，节点数量需要同时满足如下要求：

- 迁移节点数+列举节点数+1 ≤100
- 迁移节点数+列举节点数+1 ≤子网可用IP数



步骤5 单击“确认”，节点新增完成，系统开始为迁移节点安装迁移插件。添加完成后，可以在节点列表中看到新添加的节点信息。

----结束

2.9.3 计费说明

在创建迁移集群时，会产生以下费用：

- 创建Master节点、迁移节点和列举节点会产生ECS服务费用。具体计费详情请参见[ECS按需计费说明](#)或[ECS价格计算器](#)。
- 如果您使用公网迁移时，会产生NAT网关服务费用。计费详情请参见[NAT网关计费说明](#)或[NAT价格计算器](#)。
- 启用LTS服务时，会产生LTS服务费用。具体计费详情请参见[LTS计费说明](#)或[LTS价格计算器](#)。

2.9.4 集群状态说明

集群所包含的状态以及状态说明参见[表2-49](#)。

表 2-49 集群状态说明

状态	说明
创建中	正在创建集群资源
创建失败	创建集群资源失败
连接中	等待master节点上线，并开始创建和安装其他节点
健康	所有节点均处于在线状态
亚健康	部分迁移节点和列举节点在线，并且至少存在一个迁移节点和一个列举节点在线
不可用	所有迁移节点或列举节点不在线
离线	master节点离线，可能是网络中断、ECS删除等原因，请排查Vpcep Service、ECS等资源情况
升级中	升级集群版本中
升级失败	升级集群版本失败
废弃中	30天内无任务运行，废弃集群中
废弃失败	废弃集群失败
删除中	删除集群中
删除失败	删除集群失败
已废弃	集群已废弃（vpcep删除）
等待创建	等待创建集群资源
安装中	安装master节点中
安装失败	安装master节点失败
等待安装	等待安装master节点
等待升级	等待升级集群版本

状态	说明
等待删除	等待删除集群

2.9.5 资源类型和安全组配置说明

本文介绍创建迁移集群所包含的资源类型以及所有节点的安全组配置。

迁移集群包含的资源类型和介绍

资源类型	创建时可配置	是否必须	规格数量	名称	描述
ECS	是	是	<ul style="list-style-type: none"> Master节点：1台 迁移节点：根据创建集群时的参数配置创建 列举节点：根据创建集群时的参数配置创建 	“oms_cluster_ecs_” + 节点id	承载迁移插件，实现对迁移任务的管理、调度以及数据的迁移。
NAT	是	否	1个	“oms_cluster_nat_” + 集群id	公网模式下，实现列举节点、迁移节点和源端通信，实现数据的列举和迁移。
LTS	是	否	1个	主机组名称： “oms_lts_host_group_” + 集群id 主机组标识： “oms_lts_host_group_label_” + 集群id 日志组名称： “oms_lts_log_group_” + 集群id 日志流名称： “oms_lts_stream_group_” + 集群id	保存和搜索迁移集群中的日志。 创建集群时未勾选“启用LTS服务”，则不会创建。

资源类型	创建时可配置	是否必须	规格数量	名称	描述
				日志接入名称: “oms_lts_access_config_” + 集群id	
VPCEP-Service	否	是	1个	regionName + “oms-cluster” + serviceId	允许 OMS云服务与迁移集群的 Master节点通信, 实现迁移任务的管理。 配置详情请查看 VPCEP-Service 配置 。
安全组	否	是	2个	“oms_cluster_sg_master_” + 集群id	迁移集群 Master 节点的安全组, 实现与 OMS云服务以及迁移集群内部通信。 安全组配置详情请查看 Master 节点安全组配置 。

资源类型	创建时可配置	是否必须	规格数量	名称	描述
				“oms_cluster_sg_agent_” + 集群id	迁移集群列举节点和迁移节点的安全组，用于迁移集群内部通信。 安全组配置详情请查看 列举节点和迁移节点安全组配置 。

VPCEP-Service 配置

配置项	详细配置	描述
虚拟私有云	在创建集群时的网络配置中选择。	-
服务类型	接口	-
连接审批	是	确保终端节点审核后方可使用。
端口映射	22、9002	用于集群安装和任务下发、管理。
后端资源类型	云服务器	连接Master节点所在的ECS虚拟机。
连接管理	OMS云服务的VPCEP端点	允许OMS云服务与迁移集群的Master节点通信。
权限管理	OMS云服务白名单	允许OMS云服务的VPCEP端点进行连接。

Master 节点安全组配置

方向	优先级	策略	类型	协议端口	目的地址/源地址	描述
入方向	1	允许	IPv4	全部	Master节点安全组	允许安全组内的虚拟机通信
入方向	1	允许	IPv6	全部	Master节点安全组	允许安全组内的虚拟机通信
入方向	1	允许	IPv4	TCP : 22	198.19.128.0/17	云服OMS云服务和Master节点通信，实现集群安装。
入方向	1	允许	IPv4	TCP : 9002	198.19.128.0/17	云服OMS云服务和Master节点通信，实现任务下发、管理
入方向	1	允许	IPv6	全部	列举和迁移节点安全组	允许列举和迁移节点与Master通信
入方向	1	允许	IPv4	全部	列举和迁移节点安全组	允许列举和迁移节点与Master通信
出方向	1	允许	IPv6	全部	::/128	放通出方向流量，实现和源端通信
出方向	1	允许	IPv4	全部	0.0.0.0/0	放通出方向流量，实现和源端通信
出方向	1	允许	IPv6	全部	列举和迁移节点安全组	允许列举和迁移节点与Master通信
出方向	1	允许	IPv4	全部	列举和迁移节点安全组	允许列举和迁移节点与Master通信

列举节点和迁移节点安全组配置

方向	优先级	策略	类型	协议端口	目的地址/源地址	描述
入方向	1	允许	IPv6	全部	Master节点安全组	允许列举和迁移节点与Master通信

方向	优先级	策略	类型	协议端口	目的地址/源地址	描述
入方向	1	允许	IPv4	全部	Master节点安全组	允许列举和迁移节点与Master通信
入方向	1	允许	IPv4	全部	列举和迁移节点安全组	允许安全组内的虚拟机通信
入方向	1	允许	IPv6	全部	列举和迁移节点安全组	允许安全组内的虚拟机通信
出方向	1	允许	IPv4	全部	0.0.0.0/0	放通出方向流量，实现和源端通信
出方向	1	允许	IPv6	全部	::/128	放通出方向流量，实现和源端通信
出方向	1	允许	IPv6	全部	Master节点安全组	允许列举和迁移节点与Master通信
出方向	1	允许	IPv4	全部	Master节点安全组	允许列举和迁移节点与Master通信

2.10 迁移 workflow

2.10.1 workflow 配额限制说明

为确保资源的合理分配和系统的稳定运行，在创建 workflow 时，设定了以下配额限制。请在创建 workflow 前了解这些限制，以避免因超出配额而导致创建失败。

- **日创建上限：**在同一项目下，每天允许创建的迁移 workflow 数量上限为50个。
- **运行中 workflow 资源总量限制：**同一用户同时进行迁移的主机数量上限为1000台，超出此数量的主机会在 workflow 第一步暂停并处于等待状态，直至有资源释放或主机完成迁移，等待中的主机将按照创建时间顺序依次自动执行 workflow 后续步骤。
- **单个 workflow 资源限制：**单个 workflow 的资源数量不能超过100个。

2.10.2 创建主机迁移 workflow

本节介绍如何使用主机迁移标准模板创建主机迁移 workflow。

注意事项

- 建议单个 workflow 中包含的主机数量不超过100台。
- 同一用户同时进行迁移的主机数量上限为1000台，超出此数量的主机会在 workflow 第一步暂停并处于等待状态，直至有资源释放或主机完成迁移，等待中的主机将按照创建时间顺序依次自动执行 workflow 后续步骤。
- 如果您是首次使用 MgC 的主机迁移 workflow，需要您同意华为云账号进行相关的委托授权。委托权限所包含的细粒度权限请前往[委托权限说明](#)查看。

使用须知

在创建主机迁移 workflows 前，请仔细阅读并了解以下使用须知内容。

项目	须知
源端下载带宽	用于源端服务器下载迁移Agent。 <ul style="list-style-type: none"> 推荐带宽：每台机器不低于30 Mbit/s。 共用带宽建议：若多台机器共用，平均不低于50 Mbit/s。
迁移带宽	<ul style="list-style-type: none"> 重要性：影响数据传输速度和迁移时间。 推荐值与预估时长：详情请参见迁移需要多长时间?
CPU和内存要求	<ul style="list-style-type: none"> 内存：至少预留520 MB。 CPU：Linux系统预留不少于0.3核，Windows系统不低于1核。
系统兼容性列表	支持迁移的操作系统列表，请参见 兼容性列表 。
主机迁移重要声明	了解迁移过程中的重要信息和免责声明，详情请参见 主机迁移服务重要声明有哪些?
约束与限制	确保迁移过程符合服务条款和限制条件，详情请参见 主机迁移约束与限制 。
计费说明	了解迁移过程中可能产生的费用，详情请参见 计费说明 。
权限配置	使用IAM进行精细的权限管理，详情请参见 权限配置 。
网络配置及迁移端口	确保源端和目的端的网络连接和端口开放满足迁移要求，详情请参见 网络配置及端口开放 。

前提条件

- 已完成[主机资源采集](#)。
- 已将待迁移的主机资源[关联应用](#)，并对应用内的主机进行[规格评估](#)。规格评估成功后，通过[购买资源](#)，批量购买目的端。如果待迁移应用中的主机已[绑定目的端](#)，则无需进行规格评估。

操作步骤

- 步骤1** 登录[迁移中心](#)管理控制台。
- 步骤2** 在左侧导航栏选择“[实施>迁移工作流](#)”，进入[迁移工作流](#)页面。在页面左上角的当前项目下拉列表中选择[迁移项目](#)。
- 步骤3** 单击页面右上角的“[创建迁移工作流](#)”按钮，进入[选择模板](#)页面。
- 步骤4** 选择[主机迁移标准模板](#)，单击“[查看步骤详情](#)”，可以查看主机迁移模板所包含的迁移阶段和步骤，以及每个阶段和步骤的详细描述。其中，步骤类型为“[自动](#)”，表示该步由系统自动完成；步骤类型为“[手动](#)”，表示该步需要用户手动操作。单击右下角的“[开始配置](#)”按钮，进行工作流配置。

步骤5 根据表2-50，配置 workflow 参数。

表 2-50 主机迁移 workflow 配置参数说明

区域	参数	说明
workflow 详情	名称	用户自定义
	描述	
	选择资源范围	<ul style="list-style-type: none"> 按应用：迁移应用内的主机。 按迁移方案：对迁移方案中配置的主机进行迁移。
应用程序	应用名称	资源范围选择“按应用”时，选择待迁移主机所关联的应用。
迁移方案	方案名称	资源范围选择“按迁移方案”时，选择待迁移主机所在的迁移方案。创建迁移方案的方法请参考 配置主机迁移方案 。
迁移网络配置	网络类型	选择公网迁移，要求目的端服务器配置有“弹性公网IP（EIP）”，迁移时会使用目的端EIP。
		选择私网迁移，需要提前配置专线、VPN、对等连接或同VPC子网，确保源端能够通过私网访问目的端。 <ul style="list-style-type: none"> 如果源端无法访问公网，需要输入源端代理服务器的私网IP和代理软件指定的代理端口。 如果源端代理服务器无法访问公网，需要将SMS-Agent安装包放至到源端能访问下载或者借助代理能访问下载的地址。SMS-Agent安装包可以在SMS控制台下载。
目的端配置	区域	目的端所在区域，即规格评估时所购买资源的区域。
	项目	目的端所在区域的项目。
	虚拟私有云	<ul style="list-style-type: none"> 若源端IP是192.168.X.X，则推荐创建的VPC网段是192.168.0.0/16，同时创建一个子网，网段也是192.168.0.0/16。 若源端IP是172.16.X.X，则推荐创建的VPC网段是172.16.0.0/12，同时创建一个子网，网段也是172.16.0.0/12。 若源端IP是10.X.X.X，则推荐创建的VPC网段是10.0.0.0/8，同时创建一个子网，网段也是10.0.0.0/8。
	子网	子网网段与虚拟私有云网段相同。

区域	参数	说明
	安全组	<ul style="list-style-type: none"> Windows系统开放8899端口、8900端口和22端口。 Linux系统当前仅支持文件级迁移，需要开放22端口。 <p>注意</p> <ul style="list-style-type: none"> 以上端口，建议只对源端服务器开放。 防火墙开放端口与操作系统开放端口保持一致。
	开启加密	<p>选择“否”：迁移工作流在创建目的端主机步骤不会设置主机系统盘和数据盘为加密盘。</p> <p>选择“是”：迁移工作流会在创建目的端主机步骤自动设置主机系统盘和数据盘为加密盘。</p> <p>使用该功能，需要先创建委托授权EVS访问KMS。如果您有授权资格，则可直接授权。授权成功后，无需再次授权。如果权限不足，需先联系拥有Security Administrator权限的用户授权，然后再重新操作。授权成功后，设置以下参数：</p> <ul style="list-style-type: none"> 从KMS密钥中选择 在下拉菜单中选择密钥名称。您可以选择使用的密钥如下： 默认密钥：成功授权EVS访问KMS，系统会创建默认密钥“evs/default”。 自定义密钥：即您已有的密钥或者新创建密钥，创建方式请参见创建密钥。 输入KMS密钥ID 输入来自其他用户的共享密钥ID，请确保密钥位于目的端区域。共享方式请参见创建密钥授权。 <p>须知</p> <ul style="list-style-type: none"> 该功能不适用于已绑定目的端的主机。 启用该功能后，会对同一工作流内的所有新建目的端主机磁盘进行加密，且采用同一个加密密钥。 加密磁盘创建完成后不支持修改。 当前密钥共享仅支持授权给账号，不支持授权给用户。
高级设置	迁移后是否开机	<ul style="list-style-type: none"> 选择“否”，迁移完成后目的端服务器自动关机。 选择“是”，迁移完成后目的端服务器保持开机状态。

区域	参数	说明
	网络限速	<ul style="list-style-type: none"> 选择“否”，代表不限流。 选择“是”，根据要迁移的源端带宽大小及业务要求，设置带宽大小。 <p>注意 当工作流程中迁移的主机数量超过10台，为该工作流程分配的带宽资源不足以支持所有迁移任务时，为了避免迁移任务之间的带宽竞争导致部分主机迁移失败，建议您设置一个合理的网络限速，以均衡地分配可用带宽。</p>
	自动安装rsync	<ul style="list-style-type: none"> 选择“否”，不在源端安装rsync组件。 选择“是”，在源端自动安装rsync组件。 <p>注意 Linux迁移依赖rsync组件，若源端未安装rsync组件，可能会导致迁移失败。</p>
	保持IP不变	主机迁移后保持私有IP不变。使用该功能存在一定的风险，需要您自行评估并承担启用该功能带来的风险。
	启动快速模式	适用于无需增量同步的场景，可提高迁移效率。默认关闭，启用后工作流程执行完全量复制步骤后，会直接跳过增量同步步骤，继续执行后续步骤。 请根据您的需求谨慎选择。
	企业项目	选择需要迁移到的企业项目。默认为“default”。

步骤6 工作流程配置完成后，单击“**下一步：确认并创建**”按钮，进入确认并创建页面。

步骤7 确认配置信息无误后，单击“**确认创建工作流**”，弹出**启动工作流程**窗口，此时工作流程已创建完成。

- 如果您希望立即开始迁移，则单击“**确认**”按钮，立即运行工作流程。
- 如果您需要在标准工作流程中**添加阶段/步骤**，则单击“**取消**”，此时迁移工作流程状态为“**等待中**”，并未开始迁移。确定工作流程步骤后，单击“**运行**”按钮，开始迁移。

步骤8 在迁移工作流程详情页面，可以查看该主机迁移工作流程的**配置信息**和**资源迁移进度**。当资源的迁移进度完成“启动迁移Agent”这一步后，会自动在SMS控制台的迁移服务器列表创建一条迁移任务，关于MgC和SMS主机资源之间的映射关系请查看**MgC主机工作流资源与SMS迁移服务器列表之间的映射关系**。

- 将鼠标移动到主机迁移进度条，在弹窗中可以查看该主机更多迁移信息。
- 当迁移步骤（进度条）到达需要**手动确认**的步骤时，将鼠标移动到进度条，在弹窗中单击**步骤状态**后的“**确认**”，才会继续进行后续迁移步骤。
- 当迁移步骤（进度条）到达“**调整磁盘分区**”步骤时，系统会识别目的端主机是否进行了磁盘扩容。
 - 是，则需要前往**SMS控制台**，对该主机进行“调整磁盘分区”操作，方法请参考**设置目的端**中的**是否调整分区**参数配置说明。调整完成后，单击**步骤状态**后的“**确认**”继续进行后续迁移步骤。

- 否，则跳过该步骤。
- 迁移步骤中的“增量同步”为可重复步骤，在进行业务验证前，可以重复执行。
- 当迁移步骤（进度条）到达“业务割接”时，代表迁移完成。请前往目的端进行业务验证，验证无误后，手动将业务切换到目的端。切换完成后单击“确认”系统将自动进行后续“源端专项清理”和“迁移任务清理”操作。

----结束

重试/全量重试

在主机迁移 workflow 运行过程中，Edge 会为每台源端服务器启动一个 SMS-Agent 迁移进程，该进程启动后会与 SMS 控制台进行通信，并接收命令执行迁移。在工作流进行到“全量迁移”和“增量同步”两个子步骤阶段，如果源端 SMS-Agent 进程与 SMS 控制台断开连接，MgC 控制台会监控到该情况，同时迁移 workflow 会失败并出现“重试”和“全量重试”的按钮。以下是导致源端 SMS-Agent 进程与 SMS 控制台断开连接的原因，请根据具体原因选择对应的处理方法：

- **原因一：**因网络异常导致源端 SMS-Agent 进程和 SMS 控制台断开连接。
该情况通常 SMS-Agent 进程依旧存在，只需要将网络恢复，等待 SMS 控制台的迁移任务恢复连接，然后在迁移 workflow 单击“重试”即可恢复迁移。

图 2-34 重试



Linux 源端检测 SMS-Agent 进程的指令：

```
# ps -ef | grep -v grep | grep linuxmain
```

Windows 源端检测 SMS-Agent 进程的 powershell 指令：

```
# Get-Process -Name SMSAgentDeploy -ErrorAction SilentlyContinue
```

- **原因二：**源端主机重启等原因导致 SMS-Agent 进程中止，使得 SMS-Agent 与 SMS 控制台断开连接。

该情况由于 SMS-Agent 进程中止，导致迁移程序无法断点续传，需要已有的 SMS 任务重新创建。针对这种情况，MgC 提供了一键“全量重试”的功能。您只需在迁移 workflow 单击“全量重试”按钮并确认，workflow 将会直接跳到“迁移任务清理”步骤对 SMS 任务进行删除。SMS 任务清理完毕后，workflow 会再次从“启动迁移 Agent”步骤运行，并在源端重新启动 SMS-Agent 进程，创建新的 SMS 迁移任务进行迁移。

图 2-35 全量重试



注意

全量重试功能会删除旧任务，重新创建SMS迁移任务，已经迁移的数据会被覆盖。

2.10.3 创建可用区迁移 workflow

本节介绍如何使用可用区迁移标准模板创建可用区迁移 workflow。

前提条件

- 已完成[主机资源采集](#)。
- 已将待迁移的主机资源[关联应用](#)，并已对应用内的主机进行[规格评估](#)。

操作步骤

- 步骤1** 登录[迁移中心](#)管理控制台。
- 步骤2** 在左侧导航栏选择“[实施>迁移 workflow](#)”，进入[迁移 workflow](#)页面。在页面左上角的[当前项目](#)下拉列表中选择[迁移项目](#)。
- 步骤3** 单击页面右上角的“[创建迁移 workflow](#)”按钮，进入[选择模板](#)页面。
- 步骤4** 选择[可用区迁移标准模板](#)，单击“[查看步骤详情](#)”，可以查看可用区迁移模板所包含的迁移阶段和步骤，以及每个阶段和步骤的详细描述。其中，步骤类型为“[自动](#)”，表示该步由系统自动完成；步骤类型为“[手动](#)”，表示该步需要用户手动操作。单击“[开始配置](#)”按钮，进行 workflow 配置。
- 步骤5** 根据[表2-51](#)参数说明，配置 workflow 参数。

表 2-51 可用区 workflow 配置参数说明

区域	参数	说明
workflow 详情	名称	用户自定义
	描述	用户自定义

区域	参数	说明
应用程序	应用名称	选择待迁移资源所关联的应用。
迁移设置	区域	选择待迁移可用区所在Region，与创建的应用配置保持一致。
	目的可用区	选择要迁移到目的端可用区，与创建的应用配置保持一致。
	网络设置	当前只支持“保留源端配置”。
	目的虚拟机创建策略	立即创建 工作流立即开始进行源端备份，源端各自独立备份完成后，立即开始生成镜像和目的端主机。
	目的端是否关机	<ul style="list-style-type: none"> 选择“是”，生成目的端后，目的端保持关机状态。 选择“否”，生成目的端后，目的端自动开机。
	增量备份前源端是否关机	<ul style="list-style-type: none"> 选择“是”，工作流在执行增量备份前，会关闭源端主机，这样可以最大限度保证新建目的端数据与源端数据的一致性。 选择“否”，工作流在执行增量备份时，依旧保持开机状态。
	创建系统盘镜像	<ul style="list-style-type: none"> 选择“是”，将基于源端系统盘创建系统盘镜像，便于后续目的端重装系统。 选择“否”，不创建系统盘镜像。
高级设置	自动删除镜像	选择开启时，业务割接后自动删除迁移过程中产生的备份、快照及镜像文件。
	保持主网卡IP不变	选择开启时，将目的端的主网卡私有IP切换为源端主网卡私有IP，并绑定源端主网卡绑定的EIP，此时源端主网卡私有IP将设置为随机IP，并解绑EIP。以上操作不支持自动回退，需要您手动回退。

步骤6 工作流配置完成后，单击“**下一步：确认并创建**”按钮，进入确认并创建页面。

步骤7 确认配置信息无误后，单击“**确认创建工作流**”，弹出**启动工作流**窗口，此时工作流已创建完成。

- 如果您希望立即开始迁移，则单击“**确认**”按钮，立即运行工作流。
- 如果您需要在标准工作流中**添加阶段/步骤**，则单击“**取消**”，此时迁移工作流状态为“**等待中**”，并未开始迁移。确定工作流步骤后，单击“**运行**”按钮，开始迁移。

步骤8 在迁移工作流详情页面，可以查看该可用区工作流的**配置信息**和**资源迁移进度**。

- 将鼠标移动到主机迁移进度条，在弹窗中可以查看该主机更多迁移信息。

- 当迁移步骤（进度条）到达需要**手动**确认的步骤时，将鼠标移动到进度条，在弹窗中单击**步骤**状态后的“确认”，才会继续进行后续迁移步骤。

---结束

2.10.4 创建存储迁移 workflow

本节介绍如何使用存储迁移标准模板创建存储迁移 workflow。

当前支持以下区域（Region）：

- 拉美-圣地亚哥
- 拉美-圣保罗
- 土耳其-伊斯坦布尔
- 亚太-曼谷
- 亚太-新加坡
- 亚太-雅加达
- 中东-利雅得
- 华北-北京四
- 华东-上海一

约束与限制

- 关于存储迁移的约束与限制请查看[存储迁移的约束与限制有哪些？](#)
- 源端为阿里云OSS：当源端对象的存储类型是深度冷归档存储时，无法使用MgC提供的解冻归档数据功能直接迁移，需要在源端先手动解冻后再迁移。

风险提示

在创建工作流时，对于源端和目的端的同名对象，我们提供了三种覆盖策略：**不覆盖**、**直接覆盖**和**条件覆盖**。选择“**不覆盖**”策略进行数据迁移时，如果迁移任务因错误中断后重启或从暂停状态恢复，可能会造成部分数据未完全迁移，而任务显示成功的现象，影响数据的完整性。为规避这一风险，**请谨慎使用“不覆盖”策略。**

前提条件

- 已提前创建**目的端对象桶**或**弹性文件系统**。
- 已提前**创建迁移集群**。

操作步骤

- 步骤1** 登录[迁移中心](#)管理控制台。
- 步骤2** 在左侧导航栏选择“**实施>迁移 workflow**”，进入**迁移 workflow**页面。在页面左上角的**当前项目**下拉列表中选择**迁移项目**。
- 步骤3** 单击页面右上角的“**创建迁移 workflow**”按钮，进入**选择模板**页面。
- 步骤4** 选择**存储迁移标准模板**，单击“**查看步骤详情**”，可以查看存储迁移模板所包含的迁移阶段和步骤，以及每个阶段和步骤的详细描述。其中，步骤类型为“**自动**”，表示该步由系统自动完成。单击“**开始配置**”按钮，进行 workflow 配置。

步骤5 根据表2-52，配置基础信息参数。

表 2-52 基础信息参数说明

参数	说明
名称	用户自定义。
区域	下拉菜单中选择目的端所在区域。
描述	用户自定义。
集群	选择用于迁移的集群，集群中包含master节点、迁移节点和列举节点。如果没有可用集群，可以 新建集群 。 须知 master节点为系统默认创建，无需配置。

步骤6 根据表2-53、表2-54，配置源端、目的端信息。

表 2-53 源端信息参数说明

参数	说明	备注
数据站点类型	目前支持迁移的源端存储类型包括： <ul style="list-style-type: none"> ● 华为云 OBS ● 阿里云 OSS ● 百度云 BOS ● 腾讯云 COS ● 金山云 KS3 ● 七牛云 KODO ● 优刻得 US3 ● 亚马逊 S3 ● 微软 Blob ● NAS_GFS ● NAS_SMB ● NAS_NFS_V3_MOUNT ● NAS_NFS_V3_PROTOCOL ● HTTP/HTTPS数据源 	-
AK	输入源端云账号的访问密钥。	数据站点类型选择云平台时填写。
SK	输入源端云账号的私有密钥。	
桶	输入源端需要迁移的桶名。	

参数	说明	备注
对象存储终端地址	<p>输入源端桶所在区域的终端节点（Endpoint）信息。</p> <p>例如：站点类型选择阿里云OSS，源端桶所在区域为华东1（杭州），则输入“oss-cn-hangzhou.aliyuncs.com”。</p>	
桶类型	<p>根据桶的实际类型选择。桶类型可以通过查看桶的基本信息获取。</p>	<p>数据站点类型为华为云OBS时选择。</p>
appId	<p>填写腾讯云APPID。</p> <p>说明 您可以在腾讯云控制台账号信息页面查看APPID。</p>	<p>数据站点类型为腾讯云COS时填写</p>
鉴权类型	<p>鉴权类型的选择支持以下两种：</p> <ul style="list-style-type: none"> ● 连接字符串。 ● 存储账户名称+密钥。 <p>获取以上两种鉴权类型的方式请参考：如何获取微软云的访问凭证。</p>	<p>数据站点类型为微软云BLOB时配置。</p>

参数	说明	备注
列表路径	<p>输入存放列表文件的路径。列表文件存放地址必须与目的端桶处于同一区域。</p> <p>将源端待迁移的文件URL和对象名称写入列表文件（每行只能写一个URL和对象名称），列表文件需要提前创建。</p> <p>列表文件约束与限制：</p> <ul style="list-style-type: none"> 列表文件类型必须为.txt，其他文件类型不做处理，并且该文件元数据中的“ContentType”只能为：“text/plain”。 txt文件行数不超过100000行。 单个列表文件大小不能超过300 MB。 列表文件存放目录下的列表文件个数不能超过10000个。 列表文件必须是UTF-8无BOM格式编码格式。 列表文件中每行长度不要超过65535，否则会导致迁移失败。 列表文件的元数据中不能设置“ContentEncoding”，否则会导致迁移失败。 列表文件中每行使用制表符（键盘上Tab键）\t分隔URL和目的端对象名称，请勿使用空格。格式为：[URL][制表符][目的端对象名称]，其中源端对象名称如果包含中文、特殊字符必须使用URL Encode对URL编码；目的端对象名称如果包含中文、特殊字符也需要使用URL Encode编码。 列表文件中每行不要添加无效空格，否则会将空格作为对象名，导致迁移失败。 	数据站点类型选择HTTP/HTTPS数据源时填写。
文件存储服务地址	输入源端文件系统的挂载地址，格式为：IP地址+:/或IP地址+:/ xxx，例如：192.1.1.1:/或192.1.1.1:/0001	数据站点类型选择NAS_SMB/NAS_NFS_V3_MOUNT/NAS_NFS_V3_PROTOCOL时填写。
路径	输入待迁移文件路径，格式为：/文件夹。	
用户名	输入具有访问源端文件系统所有文件权限的用户名，例如：administrator。	数据站点类型选择NAS_SMB时填写。
密码	输入用户名对应的密码。	

参数	说明	备注
Windows上的域	输入对应的域名。 说明 只需输入.com之前的内容，例如：域名为test.com，则输入test。	

表 2-54 目的端信息参数说明

参数	说明	备注
数据站点类型	根据源端存储类型，选择华为云对应的存储类型。	-
AK	输入目的端华为云账号的访问密钥。	数据站点类型选择 华为云OBS 时填写。
SK	输入目的端华为云账号的私有密钥。	
桶	选择目的端OBS桶。	
对象存储终端地址	输入目的端OBS桶所在区域的终端节点（Endpoint）信息。例如：目的端桶所在区域为 华北-北京四 ，则输入“obs.cn-north-4.myhuaweicloud.com”。 说明 可在OBS桶的基本信息查看。	
指定前缀	将对象迁移到指定前缀，例如：源端对象路径为/A/B/C.txt，目的端前缀配置为/D，迁移后目的端对象路径为/D/A/B/C.txt。详细介绍请参考 设置“保存到指定前缀”参数 中的功能介绍。	
文件存储服务地址	输入目的端文件系统的挂载地址。挂载地址可以在文件系统列表页面，单击 挂载地址 后面的  图标复制获取。	数据站点类型选择 NAS_SMB/NAS_NFS_V3_MOUNT 时填写。
路径	输入目的端接收文件的路径，格式为：/文件夹。	
用户名	输入具有访问目的端文件系统所有文件权限的用户名，例如：administrator。	数据站点类型选择 NAS_SMB 时填写。
密码	输入用户名对应的密码。	
Windows上的域	输入对应的域名。 说明 只需输入.com之前的内容，例如：域名为test.com，则输入test。	

步骤7 根据表2-55，进行迁移配置。

表 2-55 迁移配置参数说明

参数	子参数	说明
迁移任务类型	全量迁移	迁移源端桶/文件夹内的全部数据。
	列表迁移	<p>通过列表文件过滤迁移，只迁移列表文件中的对象。 在“列表路径”输入存放在目的端桶的对象列表文件路径。 对象列表文件的约束与限制如下：</p> <ul style="list-style-type: none"> ● 目的端为OBS <ul style="list-style-type: none"> - 对象列表文件大小不能超过30 MB。 - 对象列表文件类型必须为.txt类型，并且元数据ContentType只能为text/plain。 - 对象列表文件必须是UTF-8无BOM格式编码格式。 - 对象列表文件每行只能包含一个对象名称，且对象名称使用URL Encode编码。 - 对象列表文件行长度不超过16KB，否则迁移失败。 - 对象列表文件元数据中不能设置ContentEncoding，否则迁移失败。 - 对象列表文件行数不超过10000行。 ● 目的端为NAS <ul style="list-style-type: none"> - 对象列表文件大小不能超过30 MB。 - 对象列表文件类型必须为.txt类型。 - 对象列表文件必须是UTF-8无BOM格式编码格式。 - 对象列表文件每行只能包含一个对象名称，且对象名称使用URL Encode编码。 - 对象列表文件行长度不超过16KB，否则迁移失败。 - 对象列表文件行数不超过10000行。
	前缀迁移	<p>仅支持云平台数据类型。 在“前缀路径”输入对象名称或者名称前缀，则只迁移与输入对象名称或名称前缀一致的对象。</p> <p>须知</p> <ul style="list-style-type: none"> ● 输入对象前缀路径时，如果对象文件在源端桶的根目录下，则直接添加前缀；如果对象文件存在目录路径，则需要对对象前缀前添加对象文件所在的目录路径。格式为：文件夹/对象前缀 ● 多个前缀之间用英文逗号分隔。
并发子任务数	-	用户自定义。配置的数目不能超过可用迁移节点数x10。例如：可用的迁移节点数为2，则最大子任务数配置数目需要≤20。

参数	子参数	说明
同名覆盖方式	不覆盖	<p>对于源端和目的端存在的同名文件，跳过源端文件，保留目的端文件。</p> <p>警告</p> <ul style="list-style-type: none"> 首次迁移时选择“不覆盖”，不会迁移源端中父文件夹的属性到目的端，可能会造成文件夹属性不完整。为规避这一风险，首次迁移时请谨慎使用“不覆盖”策略。 选择“不覆盖”策略进行数据迁移时，如果迁移任务因错误中断后重启或从暂停状态恢复，可能会造成部分数据未完全迁移，而任务显示成功的现象，影响数据的完整性。为规避这一风险，请谨慎使用“不覆盖”策略。
	直接覆盖	对于源端和目的端存在的同名文件，源端文件直接覆盖目的端文件。
	条件覆盖	<ul style="list-style-type: none"> 对于文件迁移：源端和目的端存在的同名文件，保留最后修改时间较新的文件；如果最后修改时间相同，比较文件大小，大小不同则源端覆盖目的端文件；如果最后修改时间和大小都相同，则跳过； 对于文件夹迁移：当文件迁移完成后再比对文件夹大写、时间和权限，如果文件夹大小不同/时间不相同/权限不相同，则源端覆盖目的端文件夹； <p>说明 对于空文件夹，覆盖策略与文件相同。</p>
迁移元数据	-	<p>选择是否迁移元数据。</p> <ul style="list-style-type: none"> 勾选，则迁移源端对象元数据。 不勾选，仅迁移Content-Type、Content-Encoding元数据。
清理集群	-	<p>选择工作流是否创建清理集群步骤。</p> <ul style="list-style-type: none"> 勾选，迁移工作流将自动创建清理集群步骤，您还可以继续配置是否删除NAT网关、安全组、VPCEP服务等。 不勾选，迁移工作流不创建清理集群步骤。

步骤8 （可选）根据表2-56，配置高级选项。

表 2-56 高级选项参数说明

参数	说明	备注
目的端存储类型	选择迁移到华为云OBS后，对象的存储类型。存储策略说明请参见 存储类型介绍 。	-

参数	说明	备注
KMS加密	<ul style="list-style-type: none"> 不勾选，迁移后对象状态与目的端桶级服务端加密状态保持一致。 勾选，所有对象在迁移到OBS桶后都将被加密存储。 <p>说明</p> <ul style="list-style-type: none"> 使用KMS服务端加密可能会降低迁移速率10%左右。 目的端桶所在区域支持KMS加密时，此参数才可用。 	仅目的端选择 华为云OBS 时支持配置。
解冻归档数据	<ul style="list-style-type: none"> 不勾选：如果遇到归档类型对象，则直接判定该对象迁移失败并跳过，继续迁移其余对象。 勾选：如果遇到归档类型对象，则自动解冻该对象并进行迁移。如果解冻失败，则判定该对象迁移失败并跳过，继续迁移其余对象。 <p>说明</p> <p>在数据迁移时，系统会对所有数据发起取回操作，源端云服务商收取相应的API调用费用和存储费用。</p>	-
过滤配置	通过黑、白名单以及时间段过滤迁移文件。过滤策略详细介绍请参见 过滤策略说明 。	
CDN迁移支持	<p>当默认域名不能满足您的迁移需求，比如您需要一个自定义域名以便使用CDN降低源端出口下载流量的费用，可以勾选此选项。在域名输入框中输入自定义域名，并选择传输协议。HTTP协议安全级别低于HTTPS协议，推荐使用HTTPS协议。</p> <p>源端为阿里云OSS、腾讯云COS时，如果使用了CDN鉴权，还需要选择“鉴权类型”并输入“鉴权密钥”。</p>	
SMN消息发送配置	<p>选择是否使用SMN消息通知功能，该功能可以通知您迁移任务的结果。</p> <ul style="list-style-type: none"> 不勾选，不接收SMN消息。 勾选，接收SMN消息。选择SMN主题，会向该主题的订阅者发送消息；选择消息发送语言；选择发送消息的触发条件，支持同时勾选。 	

参数	说明	备注
流量控制	<p>设置迁移任务在指定时段的最大流量带宽。</p> <ul style="list-style-type: none"> 不勾选，不进行流量控制。 勾选，进行流量控制。分别设置“开始时间”、“结束时间”和“限速”数值。 例如：设置开始时间08:00、结束时间12:00、限速为20 MB/s，当迁移任务运行到（08:00~12:00）该时间段时，最大迁移速度将限制为20 MB/s。该时间段以外，迁移速度不做限制。 <p>说明</p> <ul style="list-style-type: none"> 限速的数值范围：0~1048576 MB/s。 最多可以创建5条记录。 时区时间以目的端Region的当地时间为准。 	-
定时迁移	<p>设置迁移任务在指定时间内自动迁移。</p> <ul style="list-style-type: none"> 不勾选，手动启动/暂停迁移任务。 勾选，设置任务在指定时间范围内进行迁移，指定时间以外则暂停迁移。 例如： <ul style="list-style-type: none"> 设置开始时间08:00、结束时间12:00，则迁移任务会在每天早晨08:00至中午12:00该时间段进行迁移。该时间段以外，则迁移任务暂停。 设置开始时间12:00、结束时间08:00，则迁移任务会在每天中午12:00至次日早晨08:00该时间段进行迁移。该时间段以外，则迁移任务暂停。 	-

步骤9 workflows配置完成后，单击“**下一步：确认并创建**”按钮，进入确认并创建页面。

步骤10 确认配置信息无误后，单击“**确认创建工作流**”，弹出**启动工作流**窗口，此时工作流已创建完成。

- 如果您希望立即开始迁移，则单击“**确认**”按钮，立即运行工作流。
- 如果您需要在标准工作流中**添加阶段/步骤**，则单击“**取消**”，此时迁移工作流状态为“**等待中**”，并未开始迁移。确定工作流步骤后，单击“**运行**”按钮，开始迁移。

步骤11 在工作流详情页面，可以查看该存储迁移工作流的**配置信息**和**资源迁移进度**。还可以进行如下操作：

- 将鼠标移动到资源迁移进度条，在弹窗中可以查看该资源更多迁移信息。
- 当迁移步骤（进度条）到达需要**手动确认**的步骤时，您需要将鼠标移动到进度条上，并在弹出的窗口中单击**步骤状态**后的“**确认**”。只有完成这一操作，系统才会继续进行后续的迁移步骤。
- 在**基础信息**区域，单击集群名称后的“**管理**”，右侧弹出集群详细信息页面。在这个页面上，您可以进行如下操作：
 - 设置流量控制规则**：您可以添加、编辑或删除流量控制规则，以确保集群内的流量按照您的需求进行管理。

- **节点管理**：您可以根据实际需求，添加/删除迁移节点和列举节点，也可以对现有节点进行插件升级。

步骤12（可选）单击资源迁移进度条或进度条弹窗中的“资源迁移进度详情”，右侧弹出该资源迁移详情页面，可以查看**任务概览**和**进度详情**等信息。还可以进行如下操作：

操作	步骤
更换集群	<p>只有OMS任务状态（即 workflow 状态）为“已暂停”时，可以进行“更换集群”操作。</p>  <p>1. 在任务概览区域，单击迁移集群名称后的“更换”。</p> <p>2. 在下拉列表中选择新集群，单击“确认”，更换集群后，系统开始自动运行。</p>
修改定时迁移	<p>1. 在任务概览区域，单击定时迁移后的“修改”。</p> <p>2. 设置开始时间和结束时间，单击“确认”。</p>
修改并发子任务数	<p>1. 在进度详情区域，单击期望并发子任务数下方的“修改”，配置期望并发的子任务数。配置要求：配置的数目不能超过可用迁移节点数x10。例如，可用的迁移节点数为2，则配置数目需要≤20。</p> <p>2. 单击“确认”，完成并发子任务数修改。</p>
添加迁移限速规则	<p>1. 在迁移速度区域，单击“添加”按钮，可以增加一条限速规则，用于限制该资源在某个时间段内的迁移速度。</p> <p>须知</p> <ul style="list-style-type: none"> - 限速取值范围：1 MB~1024 GB。 - 创建多条限速规则时，限速时段不能重叠。 例如：已有一条规则，限速时段是8:00~12:00，那么7:00~13:00，7:00~8:00，9:00~12:00...有时间重叠的规则，都无法创建成功。 - 开始时间不能晚于等于结束时间。 例如：不能设置时间段为23:00~01:00。 <p>2. 设置完成后，单击“保存”，完成添加。</p>
获取迁移失败/跳过/成功资源列表文件	<p>在文件统计区域，可以查看存放迁移失败/跳过/成功的资源列表文件路径。单击路径，可以进入存放文件的OBS桶，获取对应列表文件。</p>
查看迁移流量统计	<p>在流程统计区域，可以查看该资源近1小时、近6小时、近24小时、整个周期迁移流量使用情况。</p>

----结束

过滤策略说明

过滤策略参数填写通配符规则说明和约束说明参见[表2-57](#)。

表 2-57 过滤策略参数说明

参数	说明	填写规则	约束说明
黑名单	命中黑名单规则，则文件不进行迁移/一致性对比。支持精确匹配和模糊匹配。	<ul style="list-style-type: none"> 精确匹配 全路径精确匹配，填写文件绝对路径，特殊字符用反斜杠 (\) 转义。 	<ul style="list-style-type: none"> 除{和}以外，不能出现连续的规则符号，例如：<code>***</code>、<code>*?</code>、<code>**?</code>、<code>?</code>、<code>*、?*</code>、<code>*{、}*、*}</code>、<code>?{、}*、*}</code>、<code>{、}</code>、<code>{、、*、}、、*</code>、<code>。</code> {和}之间的模式，只能使用*做通配符。 {和}之间的模式，不能有{}嵌套。 文件绝对路径在黑名单和白名单同时命中，则黑名单生效，即不做迁移。 需要配置多个匹配规则时，用英文分号隔开。
白名单	<ul style="list-style-type: none"> 白名单为空时，迁移所有文件。 白名单不为空时，仅对绝对路径命中白名单规则的文件进行迁移/一致性对比。 	<p>注意 配置黑、白名单注意事项：</p> <ul style="list-style-type: none"> 如果文件存储服务地址是以冒号加斜杠 (:/) 结尾，请确保在黑、白名单中填写的路径是相对于挂载地址的绝对路径。 例如，文件存储服务地址为 "192.1.1.1:/", 挂载地址为 "/mnt/turbo", 则填写 "/mnt/turbo" 下的绝对路径。 挂载信息示例： <pre>[root@oms-cluster-ecs filter_test]# pwd /mnt/sts_turbo/mgc/filter_test [root@oms-cluster-ecs filter_test]# ll drwxr-xr-x 2 root root 0 Aug 16 15:27 test2 -rw-r--r-- 1 root root 5 Aug 16 15:27 test2.log</pre> 如果要使黑名单精确匹配以避免迁移 "test2.log" 文件，则应在黑名单中填写路径：<code>/mgc/filter_test/test2.log</code> 如果文件存储服务地址在冒号加斜杠 (:/) 之后还包含其他路径，配置黑、白名单时，除了填写挂载地址下的绝对路径外，还要加上文件存储服务地址:/之后的部分。 例如，文件存储服务地址为 "192.1.1.1:/mgc-test", 挂载地址为 "/mnt/turbo", 则填写 "mgc-test+/mnt/turbo" 下的绝对路径。 挂载信息示例： <pre>[root@oms-cluster-ecs execution-service]# cd /mnt/turbo/autotest/filter_test/ [root@oms-cluster-ecs filter_test]# ll -rw-r--r-- 1 root root 14 Aug 8 09:22 test1.log drwxr-xr-x 1 root root 4096 Aug 8 09:22 test2 -rw-r--r-- 1 root root 14 Aug 8 09:22 test2.log</pre> 如果要使黑名单精确匹配以避免迁移 "test2" 文件，则应在黑名单中填写路径：<code>/mgc-test/autotest/filter_test/test2</code> <ul style="list-style-type: none"> 模糊匹配 	

参数	说明	填写规则	约束说明
		<ul style="list-style-type: none"> - *匹配任何字符0个或者多个，但不匹配斜杠(/) - 使用**匹配任何字符0个或者多个，包括斜杠(/) - ?匹配除斜杠(/)之外的任何字符，只能匹配一个 - {和}定义元素之间的选择：包含一个以逗号分隔的模式列表，其中任何一种模式都可能匹配，可以包含通配符。 - 特殊字符可以在它们之前使用\进行转义，\后面为通配符时，表示转义，否则表示\本身 	
时间段	<p>配置时间段，根据配置，迁移最后修改时间在配置时间段内的文件或者文件夹。</p> <p>开始时间和结束时间可以配置为空，为空时表示不限制，时间可以精确到分钟。</p>		

过滤策略场景规则示例以及举例见下表。

须知

例如第一级目录为“test”，那么：

- NAS的路径格式为：“/test”。
- OBS的路径格式为：“test”。

场景	NAS规则示例	OBS规则示例	说明
文件路径以xx结尾(特定后缀)	/xx /**xx	xx/**xx	xx可以是包含*和?的表达式
根路径下第一层xx开始的文件	/xx* <ul style="list-style-type: none"> • /testssss 匹配 • /test/xx 不匹配 	/xx* <ul style="list-style-type: none"> • testssss 匹配 • test/xx 不匹配 	
文件路径以xx开始	/xx** /xx/**	xx** xx/**	
文件名包含xx	**xx*	**xx*	
文件路径包含xx	**xx**	**xx**	

场景	NAS规则示例	OBS规则示例	说明
文件路径以xx开始，以yy结尾	/xx**yy	xx**yy	xx, yy可以是包含*和?的表达式
文件路径以xx, yy结尾(特定后缀)	**{xx,yy}	**{xx,yy}	
文件名包含xx, yy	**{xx,yy}*	**{xx,yy}*	
文件路径包含xx, yy	**{xx,yy}**	**{xx,yy}**	

2.10.5 创建对象存储批量迁移 workflow

本节介绍如何使用对象存储批量迁移模板，批量创建对象存储数据迁移任务，实现对对象存储数据的高效迁移。

约束与限制

- 关于存储迁移的约束与限制请查看[存储迁移的约束与限制有哪些?](#)
- 源端为阿里云OSS：当源端对象的存储类型是深度冷归档存储时，无法使用MgC提供的解冻归档数据功能直接迁移，需要在源端先手动解冻后再迁移。

风险提示

在创建工作流时，对于源端和目的端的同名对象，我们提供了三种覆盖策略：**不覆盖**、**直接覆盖**和**条件覆盖**。选择“**不覆盖**”策略进行数据迁移时，如果迁移任务因错误中断后重启或从暂停状态恢复，可能会造成部分数据未完全迁移，而任务显示成功的现象，影响数据的完整性。为规避这一风险，**请谨慎使用“不覆盖”策略**。

前提条件

已完成[对象存储迁移方案配置](#)。

操作步骤

- 步骤1** 登录[迁移中心](#)管理控制台。
- 步骤2** 在左侧导航栏选择“**实施>迁移工作流**”，进入**迁移工作流**页面。在页面左上角的**当前项目**下拉列表中选择**迁移项目**。
- 步骤3** 单击页面右上角的“**创建迁移工作流**”按钮，进入**选择模板**页面。
- 步骤4** 选择**对象存储批量迁移模板**，单击“**查看步骤详情**”，可以查看该模板所包含的迁移阶段和步骤，以及每个阶段和步骤的详细描述。其中，步骤类型为“**自动**”，表示该步由系统自动完成。单击“**开始配置**”按钮，进行工作流配置。
- 步骤5** 在**基础信息**区域，输入自定义的工作流**名称**和**描述**。
- 步骤6** 在**迁移方案**区域，选择**创建的对象存储迁移方案**。选择方案后，可以看到该方案的配置概览。单击“**查看详情**”，可以查看该方案的更多信息。



步骤7 在**迁移集群**区域，选择迁移使用的集群。默认使用迁移方案中配置的集群，您也可以在下拉列表中重新选择集群，选择后仅对当前工作流生效，不会修改迁移方案中配置的集群。

步骤8 根据表2-58，进行迁移配置。

表 2-58 迁移配置参数说明

参数	子参数	说明
并发子任务数	-	用户自定义。配置的数目不能超过可用迁移节点数x10。例如：可用的迁移节点数为2，则最大子任务数配置数目需要≤20。
同名覆盖方式	不覆盖	对于源端和目的端存在的同名文件，跳过源端文件，保留目的端文件。 警告 <ul style="list-style-type: none"> 首次迁移时选择“不覆盖”，不会迁移源端中父文件夹的属性到目的端，可能会造成文件夹属性不完整。为规避这一风险，首次迁移时请谨慎使用“不覆盖”策略。 选择“不覆盖”策略进行数据迁移时，如果迁移任务因错误中断后重启或从暂停状态恢复，可能会造成部分数据未完全迁移，而任务显示成功的现象，影响数据的完整性。为规避这一风险，请谨慎使用“不覆盖”策略。
	直接覆盖	对于源端和目的端存在的同名文件，源端文件直接覆盖目的端文件。
	条件覆盖	<ul style="list-style-type: none"> 对于文件迁移：源端和目的端存在的同名文件，保留最后修改时间较新的文件；如果最后修改时间相同，比较文件大小，大小不同则源端覆盖目的端文件；如果最后修改时间和大小都相同，则跳过； 对于文件夹迁移：当文件迁移完成后再比对文件夹大写、时间和权限，如果文件夹大小不同/时间不相同/权限不相同，则源端覆盖目的端文件夹； 说明 对于空文件夹，覆盖策略与文件相同。
迁移元数据	-	选择是否迁移元数据。 <ul style="list-style-type: none"> 勾选，则迁移源端对象元数据。 不勾选，仅迁移Content-Type、Content-Encoding元数据。

步骤9 （可选）根据表2-59，配置高级选项。

表 2-59 高级选项参数说明

参数	说明
目的端存储类型	选择迁移到华为云OBS后，对象的存储类型。存储策略说明请参见 存储类型介绍 。
KMS加密	<ul style="list-style-type: none"> 不勾选，迁移后对象状态与目的端桶级服务端加密状态保持一致。 勾选，所有对象在迁移到OBS桶后都将被加密存储。 <p>说明</p> <ul style="list-style-type: none"> 使用KMS服务端加密可能会降低迁移速率10%左右。 目的端桶所在区域支持KMS加密时，此参数才可用。
解冻归档数据	<ul style="list-style-type: none"> 不勾选：如果遇到归档类型对象，则直接判定该对象迁移失败并跳过，继续迁移其余对象。 勾选：如果遇到归档类型对象，则自动解冻该对象并进行迁移。如果解冻失败，则判定该对象迁移失败并跳过，继续迁移其余对象。 <p>说明</p> <p>在数据迁移时，系统会对所有数据发起取回操作，源端云服务商收取相应的API调用费用和存储费用。</p>
过滤配置	通过黑、白名单以及时间段过滤迁移文件。过滤策略详细介绍请参见 过滤策略说明 。
CDN迁移支持	<p>当默认域名不能满足您的迁移需求，比如您需要一个自定义域名以便使用CDN降低源端出口下载流量的费用，可以勾选此选项。在域名输入框中输入自定义域名，并选择传输协议。HTTP协议安全级别低于HTTPS协议，推荐使用HTTPS协议。</p> <p>源端为阿里云OSS、腾讯云COS时，如果使用了CDN鉴权，还需要选择“鉴权类型”并输入“鉴权密钥”。</p>
SMN消息发送配置	<p>选择是否使用SMN消息通知功能，该功能可以通知您迁移任务的结果。</p> <ul style="list-style-type: none"> 不勾选，不接收SMN消息。 勾选，接收SMN消息。选择SMN主题，会向该主题的订阅者发送消息；选择消息发送语言；选择发送消息的触发条件，支持同时勾选。

参数	说明
流量控制	<p>设置迁移任务在指定时段的最大流量带宽。</p> <ul style="list-style-type: none"> 不勾选，不进行流量控制。 勾选，进行流量控制。分别设置“开始时间”、“结束时间”和“限速”数值。 例如：设置开始时间08:00、结束时间12:00、限速为20 MB/s，当迁移任务运行到（08:00~12:00）该时间段时，最大迁移速度将限制为20 MB/s。该时间段以外，迁移速度不做限制。 <p>说明</p> <ul style="list-style-type: none"> 限速的数值范围：0~1048576 MB/s。 最多可以创建5条记录。 时区时间以目的端Region的当地时间为准。

步骤10 workflows配置完成后，单击“**下一步：确认并创建**”按钮，进入确认并创建页面。

步骤11 确认配置信息无误后，单击“**确认创建工作流**”，弹出**启动工作流**窗口，此时工作流已创建完成。

- 如果您希望立即开始迁移，则单击“**确认**”按钮，立即运行工作流。
- 如果您需要在 workflows 中**添加阶段/步骤**，则单击“**取消**”，此时迁移工作流状态为“**等待中**”，并未开始迁移。确定工作流步骤后，单击“**运行**”按钮，开始迁移。

步骤12 在工作流详情页面，可以查看该工作流的**配置信息**和**资源迁移进度**。还可以进行如下操作：

- 将鼠标移动到资源迁移进度条，在弹窗中可以查看该资源更多迁移信息。
- 当迁移步骤（进度条）到达需要**手动确认**的步骤时，您需要将鼠标移动到进度条上，并在弹出的窗口中单击**步骤状态**后的“**确认**”。只有完成这一操作，系统才会继续进行后续的迁移步骤。
- 在**基础信息**区域，单击集群名称后的“**管理**”，右侧弹出集群详细信息页面。在这个页面上，您可以进行如下操作：
 - 设置流量控制规则**：您可以添加、编辑或删除流量控制规则，以确保集群内的流量按照您的需求进行管理。
 - 节点管理**：您可以根据实际需求，添加/删除迁移节点和列举节点，也可以对现有节点进行插件升级。

----结束

2.10.6 创建文件存储批量迁移工作流

本节介绍如何使用文件存储批量迁移模板，批量创建文件存储数据迁移任务，实现文件存储数据的高效迁移。

约束与限制

关于文件存储迁移的约束与限制请查看[存储迁移的约束与限制有哪些？](#)

风险提示

在创建 workflows 时，对于源端和目的端的同名对象，我们提供了三种覆盖策略：**不覆盖**、**直接覆盖**和**条件覆盖**。选择“**不覆盖**”策略进行数据迁移时，如果迁移任务因错误中断后重启或从暂停状态恢复，可能会造成部分数据未完全迁移，而任务显示成功的现象，影响数据的完整性。为规避这一风险，**请谨慎使用“不覆盖”策略**。

前提条件

已完成[文件存储批量迁移方案配置](#)。

操作步骤

步骤1 登录[迁移中心管理控制台](#)。

步骤2 在左侧导航栏选择“**实施>迁移工作流**”，进入**迁移工作流**页面。在页面左上角的**当前项目**下拉列表中选择**迁移项目**。

步骤3 单击页面右上角的“**创建迁移工作流**”按钮，进入**选择模板**页面。

步骤4 选择**文件存储批量迁移模板**，单击“**查看步骤详情**”，可以查看该模板所包含的迁移阶段和步骤，以及每个阶段和步骤的详细描述。其中，步骤类型为“**自动**”，表示该步由系统自动完成。单击“**开始配置**”按钮，进行工作流配置。



步骤5 在**基础信息**区域，输入自定义的工作流**名称**和**描述**。

步骤6 在**迁移方案**区域，选择**创建的文件存储迁移方案**。选择方案后，可以看到该方案的配置概览。单击“**查看详情**”，可以查看该方案的更多信息。



步骤7 在**迁移集群**区域，选择迁移使用的**集群**。默认使用迁移方案中配置的**集群**，您也可以在下拉列表中重新选择**集群**，选择后仅对当前工作流生效，不会修改迁移方案中配置的**集群**。

步骤8 根据[表2-60](#)，进行迁移配置。

表 2-60 迁移配置参数说明

参数	子参数	说明
并发子任务数	-	用户自定义。配置的数目不能超过可用迁移节点数x10。例如：可用的迁移节点数为2，则最大子任务数配置数目需要≤20。
同名覆盖方式	不覆盖	对于源端和目的端存在的同名文件，跳过源端文件，保留目的端文件。 警告 <ul style="list-style-type: none"> 首次迁移时选择“不覆盖”，不会迁移源端中父文件夹的属性到目的端，可能会造成文件夹属性不完整。为规避这一风险，首次迁移时请谨慎使用“不覆盖”策略。 选择“不覆盖”策略进行数据迁移时，如果迁移任务因错误中断后重启或从暂停状态恢复，可能会造成部分数据未完全迁移，而任务显示成功的现象，影响数据的完整性。为规避这一风险，请谨慎使用“不覆盖”策略。
	直接覆盖	对于源端和目的端存在的同名文件，源端文件直接覆盖目的端文件。
	条件覆盖	<ul style="list-style-type: none"> 对于文件迁移：源端和目的端存在的同名文件，保留最后修改时间较新的文件；如果最后修改时间相同，比较文件大小，大小不同则源端覆盖目的端文件；如果最后修改时间和大小都相同，则跳过； 对于文件夹迁移：当文件迁移完成后再比对文件夹大写、时间和权限，如果文件夹大小不同/时间不相同/权限不相同，则源端覆盖目的端文件夹； 说明 对于空文件夹，覆盖策略与文件相同。
迁移元数据	-	选择是否迁移元数据。 <ul style="list-style-type: none"> 勾选，则迁移源端对象元数据。 不勾选，仅迁移Content-Type、Content-Encoding元数据。

步骤9 （可选）根据表2-61，配置高级选项。

表 2-61 高级选项参数说明

参数	说明
割接前是否增量同步	<ul style="list-style-type: none"> 选择是，会在迁移阶段的业务验证步骤后增加自动执行的“增量迁移”步骤。 选择否，在迁移阶段的业务验证步骤执行完后，直接进入业务割接阶段。
过滤配置	通过黑、白名单以及时间段过滤迁移文件。过滤策略详细介绍请参见 过滤策略说明 。

参数	说明
SMN消息发送配置	<p>选择是否使用SMN消息通知功能，该功能可以通知您迁移任务的结果。</p> <ul style="list-style-type: none"> 不勾选，不接收SMN消息。 勾选，接收SMN消息。选择SMN主题，会向该主题的订阅者发送消息；选择消息发送语言；选择发送消息的触发条件，支持同时勾选。
流量控制	<p>设置迁移任务在指定时段的最大流量带宽。</p> <ul style="list-style-type: none"> 不勾选，不进行流量控制。 勾选，进行流量控制。分别设置“开始时间”、“结束时间”和“限速”数值。 例如：设置开始时间08:00、结束时间12:00、限速为20 MB/s，当迁移任务运行到（08:00~12:00）该时间段时，最大迁移速度将限制为20 MB/s。该时间段以外，迁移速度不做限制。 <p>说明</p> <ul style="list-style-type: none"> 限速的数值范围：0~1048576 MB/s。 最多可以创建5条记录。 时区时间以目的端Region的当地时间为准。

步骤10 workflows配置完成后，单击“**下一步：确认并创建**”按钮，进入确认并创建页面。

步骤11 确认配置信息无误后，单击“**确认创建工作流**”，弹出**启动工作流**窗口，此时工作流已创建完成。

- 如果您希望立即开始迁移，则单击“**确认**”按钮，立即运行工作流。
- 如果您需要在 workflows 中**添加阶段/步骤**，则单击“**取消**”，此时迁移工作流状态为“**等待中**”，并未开始迁移。确定工作流步骤后，在 workflows 列表单击操作列“**运行**”按钮，开始迁移。

步骤12 单击 workflows 名称，进入 workflows 详情页面，可以查看该 workflows 的**配置信息**和**资源迁移进度**。还可以进行如下操作：

- 将鼠标移动到资源迁移进度条，在弹窗中可以查看该资源更多迁移信息。
- 当迁移步骤（进度条）到达需要**手动确认**的步骤时，您需要将鼠标移动到进度条上，并在弹出的窗口中单击**步骤状态**后的“**确认**”。只有完成这一操作，系统才会继续进行后续的迁移步骤。
- 在**基础信息**区域，单击集群名称后的“**管理**”，右侧弹出集群详细信息页面。在这个页面上，您可以进行如下操作：
 - 设置流量控制规则**：您可以添加、编辑或删除流量控制规则，以确保集群内的流量按照您的需求进行管理。
 - 节点管理**：您可以根据实际需求，添加/删除迁移节点和列举节点，也可以对现有节点进行插件升级。

----结束

2.10.7 添加阶段/步骤

您可以在 workflows 标准模板的迁移阶段/步骤，添加卡点，对迁移阶段/步骤进行控制。

注意

- 只有迁移阶段/步骤状态为**等待中/暂停**，才可以选择该阶段/步骤进行添加。迁移阶段/步骤处于**运行中/完成**状态时，无法添加。
- 如果上一阶段/步骤处于**运行中/暂停/完成**状态，该阶段/步骤只能选择**添加后置阶段/步骤**。
- **可重复**类型的步骤，无法选择**添加后置步骤**。

添加阶段

- 步骤1** 在**迁移 workflow 详情**页面，将鼠标移动到迁移阶段名称上，在弹出的窗口中，选择**添加前置阶段/添加后置阶段**，弹出**添加阶段**窗口。
- 步骤2** 为迁移阶段自定义**名称和说明**，单击“**添加步骤**”，选择步骤类型，为步骤自定义名称和说明，单击“**确认**”按钮。支持添加多个步骤。
- 步骤3** 迁移阶段中的步骤添加完成后，单击“**确认**”按钮，完成迁移阶段的添加。

须知

只有自行添加的阶段，可以进行修改、删除操作，模板中的迁移阶段无法进行修改和删除。

----结束

添加步骤

- 步骤1** 在**迁移 workflow 详情**页面，将鼠标移动到步骤名称上，选择**添加前置步骤/添加后置步骤**，弹出**添加步骤**窗口。
- 步骤2** 根据**表2-62**，选择步骤类型，为步骤自定义名称和说明，单击“**确认**”按钮，添加步骤完成。

表 2-62 步骤类型说明

步骤类型	说明
卡点	迁移到该步骤时，需要手动确认。

- 步骤3** 返回迁移阶段，查看添加的步骤。

须知

只有自行添加的步骤，可以进行修改、删除操作，模板中的步骤无法进行修改和删除。

----结束

2.11 大数据数据迁移

2.11.1 使用须知

约束与限制

- MaxCompute不支持datetime类型数据格式，在迁移之前，需要将所有datetime类型的字段转换为字符串（STRING）格式。
- MaxCompute不支持varchar和char类型字段，在迁移之前，需要将所有varchar和char类型的字段转换为字符串（STRING）格式。
- 元数据迁移不支持事务表（即DDL中带有transactional=true字段的表）。
- 由于阿里云最近更新了timestamp_ntz字段类型，目前该字段类型尚未被阿里云SDK所兼容支持，这一兼容性问题导致SDK无法识别并读取该字段，从而使该字段所在的表元数据迁移失败。

注意事项

- **CSV模板文件编辑**
当使用MgC提供的CSV模板文件进行数据筛选时，请不要使用Excel对CSV模板文件进行编辑，通过Excel编辑并保存的模板文件无法被MgC识别。
- **DLI队列连接选择**
元数据迁移和数据迁移需要使用不同的DLI队列连接，**元数据迁移**需要使用**DLI SQL队列连接**，**数据迁移**需要使用**DLI 通用队列连接**。配置错误的队列连接类型会导致迁移任务失败，因此在配置任务时，必须仔细检查并确保任务类型与队列类型相匹配。
- **元数据迁移与库的数量**
在进行元数据迁移时，MgC会根据识别到的源端数据库的数量，在DLI中自动创建相应的库。如果是整库迁移，通常只会创建一个库；而表迁移可能需要创建多个库。通常，DLI的库数量上限为10。如果DLI空间不足，可能会导致元数据迁移失败，因此在元数据迁移前，需要评估源端库的数量，确保DLI有足够的空间来创建新的库。

2.11.2 准备工作

为了确保迁移过程的顺利进行，需要完成以下必要的准备工作。

华为账号准备

使用MgC之前，您需要拥有一个华为账号或一个可用于访问MgC的IAM用户，并获取账号/IAM用户的访问密钥（AK/SK）。获取方法请参考[准备工作](#)。

阿里云访问密钥（AK/SK）准备

准备一组源端阿里云账号的Access Key（AK）和Secret Key（SK），获取方法请参考[查看RAM用户的AccessKey信息](#)。

确保该AK/SK具备以下最小权限要求：

- 拥有OSS服务的只读权限：**AliyunReadOnlyAccess**
- 拥有MaxCompute服务的只读权限：**AliyunMaxComputeReadOnlyAccess**

配置权限的方法参考[为RAM用户授权](#)。

(可选) 如果迁移的元数据**存在分区表**，需要为迁移账号开通**Information Schema**权限，方法请参考[RAM用户访问授权](#)。

创建迁移项目

在MgC控制台为本次迁移创建独立的项目，方法请参考[项目管理](#)。

配置委托权限与服务授权

为了确保正常使用DLI的功能，需要配置一条包含DLI权限和OBS权限的委托权限。

步骤1 登录华为云管理控制台。

步骤2 单击右上方登录的用户名，在下拉列表中选择“统一身份认证”。

步骤3 在左侧导航栏中，单击“委托”。

步骤4 在“委托”页面，单击“创建委托”。



步骤5 在“创建委托”页面，设置如下参数：

- 委托名称：用户自定义，例如“dli_obs_agency_access”。
- 委托类型：选择“云服务”。
- 云服务：在下拉列表中选择“数据湖探索 DLI”。
- 持续时间：按实际需求选择。
- 描述：非必选。



步骤6 配置完委托的基本信息后，单击“下一步”，进入**选择策略**页签。

步骤7 单击右上角的“新建策略”，参考**步骤8**和**步骤9**，分别创建一条OBS策略和一条DLI策略。如果已创建的授权策略中有所需的权限策略，则无需创建，直接选择已有策略即可。

步骤8 配置策略信息。

1. 策略名称：用户自定义，例如：dli-obs-agency。
2. 策略配置方式：选择“JSON视图”。
3. 在策略内容中粘贴以下自定义策略。

 **注意**

请替换“bucketName”为存放jar包的桶名称。

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "obs:bucket:GetBucketPolicy",
        "obs:bucket:GetLifecycleConfiguration",
        "obs:bucket:GetBucketLocation",
        "obs:bucket:ListBucketMultipartUploads",
        "obs:bucket:GetBucketLogging",
        "obs:object:GetObjectVersion",
        "obs:bucket:GetBucketStorage",
        "obs:bucket:GetBucketVersioning",
        "obs:object:GetObject",
        "obs:object:GetObjectVersionAcl",
        "obs:object:DeleteObject",
        "obs:object:ListMultipartUploadParts",
        "obs:bucket:HeadBucket",
        "obs:bucket:GetBucketAcl",
        "obs:bucket:GetBucketStoragePolicy",
        "obs:object:AbortMultipartUpload",
        "obs:object:DeleteObjectVersion",
        "obs:object:GetObjectAcl",
        "obs:bucket:ListBucketVersions",
        "obs:bucket:ListBucket",
        "obs:object:PutObject"
      ],
      "Resource": [
        "OBS:*:*:bucket:bucketName", //请替换bucketName为存放jar包的桶名称
        "OBS:*:*:object:*"
      ]
    },
    {
      "Effect": "Allow",
      "Action": [
        "obs:bucket:ListAllMyBuckets"
      ]
    }
  ]
}
```

步骤9 配置DLI策略信息。

1. 策略名称：用户自定义，例如：dli-agency。
2. 策略配置方式：选择“JSON视图”。

3. 在策略内容中粘贴以下自定义策略。

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "dli:table:showPartitions",
        "dli:table:alterTableAddPartition",
        "dli:table:alterTableAddColumns",
        "dli:table:alterTableRenamePartition",
        "dli:table:delete",
        "dli:column:select",
        "dli:database:dropFunction",
        "dli:table:insertOverwriteTable",
        "dli:table:describeTable",
        "dli:database:explain",
        "dli:table:insertIntoTable",
        "dli:database:createDatabase",
        "dli:table:alterView",
        "dli:table:showCreateTable",
        "dli:table:alterTableRename",
        "dli:table:compaction",
        "dli:database:displayAllDatabases",
        "dli:database:dropDatabase",
        "dli:table:truncateTable",
        "dli:table:select",
        "dli:table:alterTableDropColumns",
        "dli:table:alterTableSetProperties",
        "dli:database:displayAllTables",
        "dli:database:createFunction",
        "dli:table:alterTableChangeColumn",
        "dli:database:describeFunction",
        "dli:table:showSegments",
        "dli:database:createView",
        "dli:database:createTable",
        "dli:table:showTableProperties",
        "dli:database:showFunctions",
        "dli:database:displayDatabase",
        "dli:table:alterTableRecoverPartition",
        "dli:table:dropTable",
        "dli:table:update",
        "dli:table:alterTableDropPartition"
      ]
    }
  ]
}
```

步骤10 OBS和DLI自定义策略创建完成后，单击“**下一步**”，返回委托授权页面。

步骤11 勾选新建的OBS和DLI自定义策略，单击“**下一步**”，授权范围方案选择“**所有资源**”。

步骤12 单击“**确定**”，完成授权。授权后需等待15-30分钟才可生效。

步骤13 参考[更新委托权限](#)，将DLI委托权限进行更新。

----结束

创建虚拟私有云 VPC

购买ECS时，需要进行网络配置，可提前创建VPC和子网，方法请参考[创建虚拟私有云和子网](#)。

 **注意**

ECS的VPC网段不能和DLI队列（弹性资源池）使用的网段冲突，DLI队列默认使用的网段为172.16.0.0/18。

购买 ECS

- **购买ECS**：在与目的端DLI同一区域（Region）下购买一台Linux系统的ECS主机，购买方法请参考[购买弹性云服务器](#)。其中网络配置选择上一步[创建的虚拟私有云](#)和[子网](#)。该ECS主机还需要满足以下条件：
 - 可以访问外网（公网），并检查是否能够访问MgC和IoTDA服务的域名，具体需要检查的域名请查看[域名列表](#)。
 - 安全组的出方向规则中，需要包含8883端口。
 - 操作系统为：CentOS 8.x。
 - 推荐规格不小于8U16G。
- **创建并绑定EIP**：ECS主机要实现访问外网，需要配置EIP（弹性公网IP）。如果主机已绑定EIP则跳过该操作无需重复绑定；如果未绑定EIP请参考[申请EIP](#)和[将EIP绑定至ECS](#)进行绑定。
 - 计费模式：建议选择[按需计费](#)。
 - 带宽大小：推荐5 Mbit/s。

安装迁移工具 Edge 并连接迁移中心

- **安装Edge工具**：在购买的ECS主机上，安装用于配合数据校验的工具Edge，方法请参见[安装Linux版本的Edge工具](#)。
- **注册Edge用户**：在浏览器中访问Linux主机的网卡IP地址+端口（27080），如：<https://x.x.x.x:27080>。首次访问后会进入注册页面，分别输入用户名、密码，确认密码无误后，单击《[隐私政策声明](#)》，仔细阅读后，勾选“我已阅读并同意”，最后单击“确认”即可完成注册。
- **连接迁移中心**：安装并注册成功后，将Edge与迁移中心进行连接，连接方法请参考[连接迁移中心](#)。

 **注意**

迁移工具Edge不具备自动重启功能，严禁在任务执行期间重启Edge，以避免任务失败。

添加凭证

在迁移工具Edge上分别添加一条MaxCompute凭证（源端阿里云账号的AK/SK）和数据湖DLI凭证（目的端华为云账号的AK/SK），认证方式选择AK/SK，添加方法请参考[添加资源凭证](#)。

- MaxCompute凭证输入源端阿里云账号的AK/SK。

创建凭证

* 使用类型

* 类型

* 凭证名称

* 认证方式

* ak

* sk

- 数据湖DLI凭证输入目的端华为云账号的AK/SK。

创建凭证

* 使用类型

* 类型

* 凭证名称

* 认证方式

* ak

* sk

创建 OBS 桶并上传 jar 包

在华为云OBS创建一个存储桶，并将数据迁移过程中所依赖的 Java 文件（jar包）上传至该OBS桶。创建OBS桶的方法请参考[创建桶](#)。上传文件的方法请参考[上传对象](#)。

数据迁移所依赖的jar包分别为：migration-dli-spark-1.0.0.jar、fastjson-1.2.54.jar、datasource.jar。这三个jar包是数据迁移过程中不可或缺的，三个jar包的用途和获取方法如下：

- **migration-dli-spark-1.0.0.jar**
 - 用途：用于创建spark会话并提交sql语句。
 - 获取方法：在[迁移工具Edge所部署主机](#)的“/opt/cloud/Edge/tools/plugins/collectors/bigdata-migration/dliSpark”路径下获取。
- **fastjson-1.2.54.jar**：
 - 用途：用于处理JSON格式的数据交换。
 - 获取方法：在[迁移工具Edge所部署主机](#)的“/opt/cloud/Edge/tools/plugins/collectors/bigdata-migration/deltaSpark”路径下获取。

- **datasource.jar:**
 - 用途：包含数据源的配置和连接逻辑，允许服务连接到不同的数据库或数据存储系统。
 - 获取方法：需要您根据需求自行获取并编译。

购买弹性资源池并添加队列

弹性资源池为DLI作业运行提供所需的计算资源（CPU和内存）。购买方法请参考[创建弹性资源池](#)。

创建完弹性资源池后，需要为弹性资源池分别添加一条**通用队列**和一条**SQL队列**，用于后续作业的运行。添加方法请参考[添加队列](#)。

创建 DLI 增强型跨源连接

步骤1 登录DLI管理控制台，在左侧导航栏单击“资源管理 > 弹性资源池”，可进入弹性资源池管理页面。

步骤2 单击弹性资源池名称前的 ，在展开的信息中，获取该弹性资源池的网段。

步骤3 将弹性资源池的网段加入ECS安全组的入方向规则中。

1. 登录弹性云服务器控制台。
2. 在弹性云服务器列表中，单击购买的Linux主机名称（用于安装Edge的主机）。
3. 选择**安全组**页签，单击安全组的“配置规则”。



4. 选择**入方向规则**，单击“添加规则”。



5. 设置入方向规则参数。
 - 优先级：设置为1。
 - 策略：选择“允许”。
 - 类型：选择“IPv4”。
 - 协议端口：选择“全部协议”。
 - 源地址：选择IP地址，并填写弹性资源池的网段。
6. 单击“确定”完成安全组配置。

步骤4 在DLI管理控制台的左侧导航栏单击“跨源管理”，进入跨源管理页面。

步骤5 选择“增强型跨源”，单击“创建”，弹出创建连接窗口。



步骤6 根据表2-63，配置增强型跨源连接信息。



表 2-63 参数配置说明

参数	配置说明
连接名称	用户自定义
弹性资源池	选择 购买的弹性资源池 。
虚拟私有云	选择 创建的虚拟私有云和子网 。
子网	
路由表	保持默认。

参数	配置说明
主机信息	<p>需要添加两条源端MaxCompute侧主机信息，填写内容和格式如下：</p> <p>EndpointIP Endpoint（中间有空格）</p> <p>TunnelEndpointIP TunnelEndpoint（中间有空格）</p> <p>两条主机信息之间用回车进行隔开。举例：</p> <p>118.178.xxx.xx service.cn- hangzhou.maxcompute.aliyun.com.vipgds.alibabadns.com</p> <p>47.97.xxx.xx dt.cn-hangzhou.maxcompute.aliyun.com</p> <p>EndpointIP和TunnelEndpointIP的获取方法：使用任一有公网IP的主机，分别对源端MaxCompute所在地域的Endpoint和Tunnel Endpoint执行ping指令，在回显中即可获取对应IP地址。</p> <p>Endpoint和Tunnel Endpoint信息请查看各地域Endpoint对照表。</p>

步骤7 单击“确定”，创建增强型跨源连接。创建完成后，增强型跨源连接的连接状态显示“已激活”，代表该连接创建成功。

----结束

添加并配置路由

- **添加路由**

为**创建的DLI增强型跨源连接**添加**两条**路由，方法请参考[添加路由](#)。路由IP与创建跨源连接时，配置的主机信息IP保持一致。

- **配置路由**

1. 登录虚拟私有云管理控制台。
2. 在左侧导航栏，选择“虚拟私有云 > 路由表”，进入路由表列表页面。
3. 在路由表列表中，找到创建跨源连接时对应的路由表（即ECS所在VPC的路由表），单击路由表名称，进入路由表详情页面。
4. 单击“添加路由”，弹出添加路由窗口。



5. 按照提示配置参数。需要添加两条路由，单击  **继续添加**，可以依次增加多条路由。
 - 目的地址类型：选择“IP地址”。
 - 目的地址：与创建跨源连接时，配置的主机信息IP保持一致。
 - 下一跳类型：选择“服务器实例”。

- 下一跳：选择**购买的ECS**。



配置 SNAT 转换规则

配置完SNAT转换规则后，如果ECS重启，则需要重新配置。

步骤1 登录**购买的ECS主机**。

步骤2 依次执行下列两条命令：

```
sysctl net.ipv4.ip_forward=1
```

该命令用于启用 Linux 主机上的 IP 转发功能。

```
iptables -t nat -A POSTROUTING -o eth0 -s {DLI弹性资源池所在网段} -j SNAT --to {ECS私网IP}
```

设置iptables规则以进行网络地址转换。

----结束

测试 DLI 队列与数据源的连通性

步骤1 登录DLI管理控制台，在左侧导航栏单击“资源管理 > 队列管理”，可进入队列管理页面。

步骤2 找到为弹性资源池**添加的DLI队列**，单击操作列的“更多 > 测试地址连通性”，弹出测试地址连通性窗口。

步骤3 参考窗口中的样例，依次输入MaxCompute侧的“Endpoint域名”、“Endpoint IP”、“Tunnel Endpoint域名”和“Tunnel Endpoint IP”进行四次连通性测试，端口建议输入443端口。



步骤4 单击“测试”。

如果测试地址可连通，页面上将提示地址可达。

如果测试地址不可连通，页面上将提示地址不可达，请检查网络配置后重试。检查网络配置即检查所测试的VPC对等连接或跨源连接是否处于已激活状态。

----结束

开通 DLI 服务 Spark 3.3 和 jar 程序访问元数据白名单

数据迁移需要使用DLI服务的Spark 3.3特性和jar程序访问DLI元数据特性，请联系DLI服务的技术支持人员申请开通。

（可选）开通 DLI 服务 Spark 生命周期白名单

如果待迁移的元数据有生命周期表（即DDL带有LIFECYCLE字段），请联系DLI服务的技术支持人员申请开通Spark生命周期特性白名单。

（可选）开通 100 网段白名单

如果您是使用**专线**迁移方式进行数据迁移，需要提前向VPC服务申请开通“100.100.x.x”网段的白名单。

请提交工单到VPC服务，并提供以下信息进行开通：

- 华为云**账号名**、目的端DLI服务所在Region的**项目ID**：获取方法请参考[API凭证](#)。
- DLI服务的**租户名**以及租户的**项目ID**：请联系DLI服务的技术支持人员获取。

2.11.3 创建源端连接

创建源端连接是为了允许迁移工具Edge访问和读取源端阿里云 MaxCompute中存储的数据。

前提条件

已完成[准备工作](#)。

操作步骤

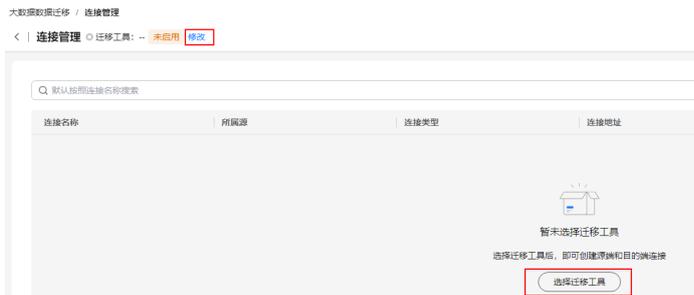
步骤1 登录[迁移中心控制台](#)。

步骤2 在左侧导航栏选择“**实施>大数据数据迁移**”，进入**大数据数据迁移**页面。在页面左上角的**当前项目**下拉列表中选择在完成[准备工作](#)时所创建的迁移项目。

步骤3 单击页面右上角的“**连接管理**”按钮，进入**连接管理**页面。



步骤4 首次使用大数据数据迁移功能时，需要先启用迁移工具Edge。单击页面左上角的“**修改**”或单击“**选择迁移工具**”按钮，弹出选择迁移工具窗口。在下拉列表中，选择在完成[准备工作](#)时与迁移中心连接成功的Edge。



注意

在进行大数据数据迁移时，请确保迁移工具Edge一直处于在线并启用状态。

步骤5 启用迁移工具Edge后，单击页面右上角的“**创建源端连接**”按钮，右侧弹出创建连接窗口。



步骤6 连接类型选择“**MaxCompute**”，单击“**下一步**”。



步骤7 根据[MaxCompute连接参数配置说明](#)，配置连接参数后，单击“**测试连接**”按钮，测试通过代表连接成功。

表 2-64 MaxCompute 连接参数配置说明

参数	配置说明
所属源	选择源端。
连接名称	默认创建名称为“MaxCompute-4位随机数（字母和数字）”，也可以自定义名称。
迁移工具	选择在完成准备工作时与迁移中心连接成功的Edge。
阿里云凭证	选择完成准备工作时，在Edge侧添加的阿里云（AK/SK）凭证。
项目名称	输入待迁移数据所在的MaxCompute项目名称。项目名称可在MaxCompute控制台的项目管理中获取。

参数	配置说明
Endpoint	输入MaxCompute服务所在地域的VPC网络Endpoint信息。 各地域VPC网络Endpoint信息请参考 Endpoint对照表 。

创建源端连接完成后，在[连接管理](#)页面，可以查看已创建的连接列表。如有需要，可以单击列表操作列的“[修改链接](#)”对连接配置进行修改。



----结束

2.11.4 创建目的端连接

创建目的端连接是为了将源端数据写入到华为云DLI。

使用须知

元数据迁移：只需要配置一条连接到DLI SQL队列的连接。

数据迁移：需要配置两条连接，分别为连接到DLI SQL队列的连接和连接到DLI通用队列的连接。

前提条件

已完成[准备工作](#)。

操作步骤

步骤1 登录[迁移中心控制台](#)。

步骤2 在左侧导航栏选择“[实施>大数据数据迁移](#)”，进入[大数据数据迁移](#)页面。在页面左上角的当前项目下拉列表中选择在完成[准备工作](#)时所创建的迁移项目。

步骤3 单击页面右上角的“[连接管理](#)”按钮，进入[连接管理](#)页面。



步骤4 单击页面右上角的“[创建目的端连接](#)”按钮，右侧弹出[新增连接](#)窗口。



步骤5 连接类型选择“数据湖探索（DLI）”，单击“下一步”，进入详细配置页面。



步骤6 根据[DLI服务参数配置说明](#)，配置连接参数后，单击“测试连接”按钮，测试通过代表连接成功。

表 2-65 DLI 服务参数配置说明

参数	配置说明
所属源	选择目的端。
连接名称	默认创建名称为“DLI-4位随机数（字母和数字）”，也可以自定义名称。
DLI凭证	选择完成 准备工作 时，在Edge侧添加的华为云（AK/SK）凭证。如果所选凭证与登录迁移中心账号凭证相同，请勾选“ 所选凭证与登录迁移中心凭证相同 ”，系统将根据所选的区域，自动查询账号在该区域下的项目。
区域	选择目的端DLI所在区域。如果未勾选“ 所选凭证与登录迁移中心凭证相同 ”，则需要手动填写区域代号。例如：cn-south-1
项目	选择DLI在目的端区域用于接收数据的项目。如果未勾选“ 所选凭证与登录迁移中心凭证相同 ”，则需要手动填写项目ID。项目ID的获取方法请参考 获取项目的名称和ID 。 
队列名称	输入完成 准备工作 时，创建的SQL队列名称/通用队列名称。 <ul style="list-style-type: none"> 当配置SQL队列连接时，填写SQL队列名称。 当配置通用队列连接时，填写通用队列名称。

创建目的端连接完成后，在[连接管理](#)页面，可以查看已创建的连接列表。如有需要，可以单击列表操作列的“[修改链接](#)”对连接配置进行修改。



----结束

2.11.5 迁移实施

2.11.5.1 元数据全量迁移

将源端数据库中的所有元数据迁移到华为云数据湖探索（DLI）。

须知

元数据迁移不支持事务表（即DDL中带有transactional=true字段的表）。

前提条件

- 已完成所有准备工作。
- 已创建源端连接。
- 已创建目的端连接。

操作步骤

步骤1 登录迁移中心控制台。

步骤2 在左侧导航栏选择“实施>大数据数据迁移”，进入大数据数据迁移页面。在页面左上角的当前项目下拉列表中选择在完成准备工作时所创建的迁移项目。

步骤3 单击页面右上角的“创建迁移任务”按钮，右侧弹出创建迁移任务窗口。



步骤4 源端组件选择“MaxCompute”，目的端组件选择“数据湖探索 DLI”，任务类型选择“元数据全量迁移”，单击“下一步”，进入详细配置页面。



步骤5 根据表2-66，配置元数据全量迁移任务参数。

表 2-66 元数据全量迁移参数配置说明

区域	参数	配置说明
基本信息	任务名称	默认创建名称为“MaxCompute到DLI元数据全量迁移-4位随机数（字母和数字）”，也可以自定义名称。
	迁移工具	选择在完成准备工作时与迁移中心连接成功的Edge。
源端配置	源端连接	选择创建的源端连接。
	MaxCompute参数	非必填参数，一般为空。如有需要，可参考MaxCompute参数文档进行配置。
数据范围	按库筛选	<p>即按数据库进行筛选，在数据库参数框中输入待迁移的数据库名称。单击“添加数据库”，可以新增数据库，最多支持同时添加10个数据库。</p> <p>数据范围</p> <p>筛选方式</p> <p><input type="button" value="按库筛选"/> <input type="button" value="按表筛选"/></p> <p>数据库</p> <p>点击“添加数据库”后，在输入框中输入数据库名</p> <p><input type="text"/></p> <p><input type="text"/></p> <p><input type="button" value="添加数据库"/></p> <p>您还可以增加9个数据库。</p> <p>企业项目ID (可选)</p> <p><input type="text" value="请输入数据库所属企业项目"/></p> <p>排除表 (可选)</p> <p><input type="button" value="添加文件"/> <input type="button" value="下载模板"/></p> <p>只能上传CSV文件，且不能超过1M，表格数据将在上传后自动去重。</p> <p>如果数据库中有不需要迁移的表，可以下载CSV格式的“排除表”模板，参考b~e.在模板中添加不需要迁移的表信息，然后上传至MgC。</p>

区域	参数	配置说明
	按表筛选	<ol style="list-style-type: none"> 1. 下载CSV格式的模板。 2. 下载后用记事本打开该CSV模板文件。 注意 请不要使用Excel对CSV模板文件进行编辑，通过Excel编辑并保存的模板文件无法被MgC识别。 3. 保留CSV文件中的第一行不变，从第二行开始，每行填写一个表的迁移信息，格式为“{MC项目名};{表名}”，其中MC项目名代表MaxCompute项目名称，表名代表具体的表名称。 须知 <ul style="list-style-type: none"> - 每行的“MC项目名”和“表名”之间使用英文逗号隔开，不要使用空格或其他分隔符。 - 每添加完一个表的信息后，使用回车键换行。 4. 完成所有表信息的添加后，保存CSV文件的更改。 5. 将编辑并保存好的CSV文件上传至MgC。
	企业项目ID（可选）	输入目的端DLI数据库所属的企业项目ID。企业项目ID的获取方法请参考 如何获取企业项目ID 。
	内表转换	<ul style="list-style-type: none"> • 选择内表进行迁移时，迁移后的数据会存储在DLI中。 • 选择外表进行迁移时，迁移后的数据会存储在您指定的OBS路径下。
目的端配置	覆盖策略	对已有表的处理策略 <ul style="list-style-type: none"> • 跳过：如果目的端已经存在同名的数据表，将忽略该表，不会对其进行任何操作。 • 重建：如果目的端已经存在同名的数据表，该表及其数据将被完全删除，并根据源端的定义重新创建。
	目的端连接	选择创建的 目的端DLI SQL队列连接 。 注意 切勿选择通用队列。
	自定义参数	根据实际需求进行配置，支持的自定义参数详情请参考 conf参数说明 和 自定义参数说明 。
	表类型	根据实际情况选择。
	文件类型	根据实际情况选择。
	压缩格式	根据实际情况选择。
迁移选项	并发连接数	设置迁移并发数。默认为3，1≤取值范围≤10。

区域	参数	配置说明
	表路径映射（可选）	<p>用于确保在元数据迁移过程中，MaxCompute外表的元数据能够正确地迁移到目的端指定的存储位置，即在OSS中的源路径自动映射为目的端指定的OBS路径。</p> <p>映射到目的端OBS的路径需要提前创建，并且除桶名以外，其余目录需要与源端OSS目录保存一致。</p> <p>例如，源端表所在目录路径为：<code>oss://oss-cn-hangzhou.aliyuncs.com/ali-test/documents/example_table</code>，其中<code>oss-cn-hangzhou.aliyuncs.com</code>为桶所在地域的Endpoint，<code>ali-test</code>为桶名，<code>documents</code>为表的根目录，<code>example_table</code>为表目录。那么，映射到目的端的路径必须为：<code>obs://hw-test/documents/example_table</code>，除桶名外，根目录和表目录必须和源端一致。</p> <p>填写示例：</p> <ul style="list-style-type: none"> 源端：<code>oss://oss-cn-hangzhou.aliyuncs.com/ali-test/documents/example_table</code> 目的端：<code>obs://hw-test/documents/example_table</code> <p>说明 以上路径仅为示例，请在实际使用中替换为正确的路径。</p>

步骤6 配置完成后，选择任务的运行方式。

须知

- 创建的迁移任务可重复运行，每运行一次，即会生成一条任务实例。
- 单击任务名称，可以修改任务配置。
- **立即运行**：如果需要立即运行迁移任务，可以勾选“**保存后立即运行**”，然后单击“**保存**”按钮，会在**任务管理**列表中生成一条元数据全量迁移任务并自动开始执行迁移。

The screenshot shows a configuration window titled "迁移选项" (Migration Options). It includes a "并发连接数" (Parallel Connections) field with a value of 3 and a note: "请输入区间为1-10的并发连接数" (Please enter a parallel connection count in the range of 1-10). Below is the "表路径映射 (可选)" (Table Path Mapping) section, which includes a note: "请补充源端表路径到目的端表路径的映射关系, 该映射关系将会在目的端创建的DDL中替换掉相应表路径字段。" (Please supplement the mapping relationship from source table path to destination table path, this mapping relationship will be used to replace the corresponding table path field in the DDL created at the destination.) There are two input fields: "请输入oss桶路径" (Please enter OSS bucket path) and "请输入obs桶路径" (Please enter OBS bucket path). At the bottom, there are three buttons: "保存后立即运行" (Save and run immediately), "取消" (Cancel), and "保存" (Save). The "保存后立即运行" checkbox is checked, and the "保存" button is highlighted with a red box.

- **手动运行**：如果需要后续手动启动迁移任务，则只需单击“**保存**”按钮，在**任务管理**列表中生成一条元数据全量迁移任务。在需要启动迁移时，单击操作列的“**运行**”，手动启动迁移。



步骤7 迁移任务运行后，单击操作列的“**查看实例**”，进入**任务进度**页签，可以查看该任务产生的运行实例以及迁移进度。



单击进度统计列的“**查看**”，进入进度详情页面，可以**查看并导出任务结果**。

----结束

2.11.5.2 元数据增量感知

识别出上一次迁移后，在源端数据库中发生变化的元数据，并同步至目的端华为云数据湖探索（DLI）。

前提条件

- 已**创建源端连接**。
- 已**创建目的端连接**。
- 最少完成一次**元数据全量迁移**。

准备工作

- **开通spark 3.3.1白名单**
使用该功能前，请联系技术支持人员，将您添加到spark 3.3.1版本的白名单中。
- **配置DLI作业桶**
需要购买一个OBS桶或并行文件系统并设置为DLI作业桶，该桶用于存储DLI作业运行过程中产生的临时数据。方法请参考**配置DLI作业桶**。

操作步骤

- 步骤1** 登录**迁移中心控制台**。
- 步骤2** 在左侧导航栏选择“**实施>大数据数据迁移**”，进入**大数据数据迁移**页面。在页面左上角的**当前项目**下拉列表中选择在完成**准备工作**时所创建的迁移项目。
- 步骤3** 单击页面右上角的“**创建迁移任务**”按钮，右侧弹出**创建迁移任务**窗口。



步骤4 源端组件选择“MaxCompute”，目的端组件选择“数据湖探索 DLI”，任务类型选择“元数据增量感知”，单击“下一步”，进入详细配置页面。



步骤5 根据表2-67，配置元数据增量感知任务参数。

表 2-67 元数据增量感知参数配置说明

区域	参数	配置说明
基本信息	任务名称	默认创建名称为“MaxCompute到DLI元数据增量感知-4位随机数（字母和数字）”，也可以自定义名称。
	迁移工具	选择在完成准备工作时与迁移中心连接成功的Edge。
源端配置	源端连接	选择创建的源端连接。
	时间范围	<ul style="list-style-type: none"> 全量区间：全量识别出上一次元数据迁移后，在源端数据库中发生变化的元数据。 增量区间：选择T-N区间作为增量元数据的识别范围，即任务开始执行时间的前 24 x N 小时内的增量元数据。例如：选择T-1，2024/06/06 14:50执行任务，则表示识别从2024/06/05 14:50 到 2024/06/06 14:50（现在）的增量元数据。
	MaxCompute参数（可选）	非必填参数，一般为空。如有需要可参考MaxCompute参数文档进行配置。

区域	参数	配置说明
数据范围	按库筛选	<p>即按数据库进行筛选，在数据库参数框中输入需要增量待迁移的数据库名称。单击“添加数据库”，可以新增数据库，最多支持同时添加10个数据库。</p>  <p>如果数据库中有不需要迁移的表，可以下载CSV格式的“排除表”模板，参考b.~e.在模板中添加不需要迁移的表信息，然后上传至MgC。</p>
	按表筛选	<ol style="list-style-type: none"> 1. 下载CSV格式的模板。 2. 下载后用记事本打开该CSV模板文件。 <p>注意 请不要使用Excel对CSV模板文件进行编辑，通过Excel编辑并保存的模板文件无法被MgC识别。</p> <ol style="list-style-type: none"> 3. 保留CSV文件中的第一行不变，从第二行开始，每行填写一个表的迁移信息，格式为“{MC项目名},{表名}”，其中MC项目名代表MaxCompute项目名称，表名代表具体的表名称。 <p>须知</p> <ul style="list-style-type: none"> - 每行的“MC项目名”和“表名”之间使用英文逗号隔开，不要使用空格或其他分隔符。 - 每添加完一个表的信息后，使用回车键换行。 <ol style="list-style-type: none"> 4. 完成所有表信息的添加后，保存CSV文件的更改。 5. 将编辑并保存好的CSV文件上传至MgC。
目的端配置	目的端连接	<p>选择创建的目的端DLI SQL队列连接。</p> <p>注意 切勿选择通用队列。</p>
	自定义参数（可选）	<p>根据实际需求进行配置，支持的自定义参数详情请参考conf参数说明和自定义参数说明。</p>
迁移选项	并发连接数	<p>设置迁移并发数。默认为3，1≤取值范围≤10。</p>

步骤6 配置完成后，选择任务的运行方式。

须知

- 创建的迁移任务可重复运行，每运行一次，即会生成一条任务实例。
 - 单击任务名称，可以修改任务配置。
- **立即运行**：如果需要立即运行迁移任务，可以勾选“**保存后立即运行**”，然后单击“**保存**”按钮，会在**任务管理**列表中生成一条元数据增量感知任务并自动开始执行迁移。



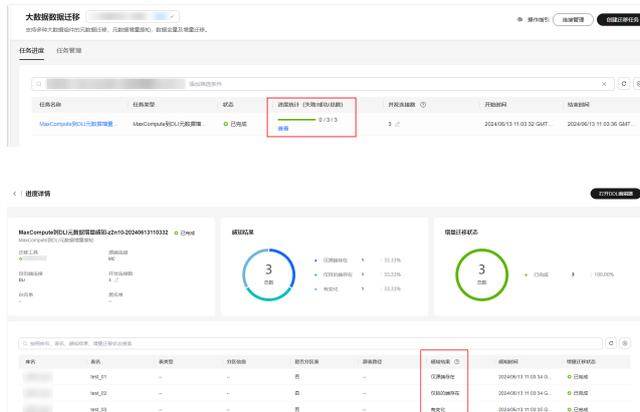
- **手动运行**：如果需要后续手动启动迁移任务，则只需单击“**保存**”按钮，在**任务管理**列表中生成一条元数据增量感知任务。在需要启动迁移时，单击操作列的“**运行**”，手动启动迁移。



步骤7 迁移任务运行后，单击操作列的“**查看实例**”，进入**任务进度**页签，可以查看该任务产生的运行实例以及迁移进度。



单击进度统计列的“**查看**”，进入**进度详情**页面。可以查看元数据的增量感知结果并**导出任务结果**。



步骤8 单击**进度详情**页面右上角的“**打开DDL编辑器**”，可以根据不同的增量感知结果，对源端和目的端的表结构进行比较和编辑。



----结束

2.11.5.3 全量数据迁移

将源端数据库中的所有数据迁移到目的端华为云数据湖探索（DLI）。

前提条件

- 已完成所有准备工作。
- 已创建源端连接。
- 已创建目的端连接。
- 开始迁移前，需要联系技术支持人员，将您添加到允许jar程序访问DLI元数据功能的白名单中。

操作步骤

步骤1 登录迁移中心控制台。

步骤2 在左侧导航栏选择“实施>大数据数据迁移”，进入大数据数据迁移页面。在页面左上角的当前项目下拉列表中选择在完成准备工作时所创建的迁移项目。

步骤3 单击页面右上角的“创建迁移任务”按钮，右侧弹出创建迁移任务窗口。



步骤4 源端组件选择“MaxCompute”，目的端组件选择“数据湖探索 DLI”，任务类型选择“全量数据迁移”，单击“下一步”，进入详细配置页面。



步骤5 根据表2-68，配置全量数据迁移任务参数。

表 2-68 全量数据迁移参数配置说明

区域	参数	配置说明
基本信息	任务名称	默认创建名称为“MaxCompute到DLI全量数据迁移-4位随机数（字母和数字）”，也可以自定义名称。
	迁移工具	选择在完成准备工作时与迁移中心连接成功的Edge。
源端配置	源端连接	选择创建的源端连接。
	预估项目周期（天）（可选）	配置后，系统会在任务执行时检查表的生命周期。若表的生命周期在项目的预期结束时间之前终止，则忽略该表。不配置则默认迁移所有表。
	MaxCompute参数（可选）	非必填参数，一般为空。如有需要，可参考 MaxCompute参数文档 进行配置。
数据范围	按库筛选	<p>即按数据库进行筛选，在数据库参数框中输入待迁移的数据库名称。单击“添加数据库”，可以新增数据库，最多支持同时添加10个数据库。</p>  <p>如果数据库中有不需要迁移的表，可以下载CSV格式的“排除表”模板，参考b.~e.在模板中添加不需要迁移的表信息，然后上传至MgC。</p>

区域	参数	配置说明
	按表筛选	<ol style="list-style-type: none"> 1. 下载CSV格式的模板。 2. 下载后用记事本打开该CSV模板文件。 注意 请不要使用Excel对CSV模板文件进行编辑，通过Excel编辑并保存的模板文件无法被MgC识别。 3. 保留CSV文件中的第一行不变，从第二行开始，每行填写一个表的迁移信息，格式为“{MC项目名},{表名}”，其中MC项目名代表MaxCompute项目名称，表名代表具体的表名称。 须知 <ul style="list-style-type: none"> - 每行的“MC项目名”和“表名”之间使用英文逗号隔开，不要使用空格或其他分隔符。 - 每添加完一个表的信息后，使用回车键换行。 4. 完成所有表信息的添加后，保存CSV文件的更改。 5. 将编辑并保存好的CSV文件上传至MgC。
目的端配置	目的端连接	<p>选择创建的目的端DLI 通用队列连接。</p> <p>注意 切勿选择DLI SQL队列。</p>

区域	参数	配置说明																																				
	自定义参数（可选）	<p>根据实际需求进行配置，支持的自定义参数详情请参考conf参数说明和自定义参数说明。</p> <ul style="list-style-type: none"> 如果是公网迁移，需要填写以下4个参数： <div data-bbox="858 421 1372 631"> <p>自定义参数 参数详情请参考文档中“表3 conf参数说明”部分。 查看文档</p> <table border="1"> <tr> <td>spark.dli.metaAccess.enable</td> <td>true</td> <td>🗑</td> </tr> <tr> <td>spark.dli.job.agency.name</td> <td></td> <td>🗑</td> </tr> <tr> <td>mgc.mc2dli.data.migration.dli.file.path</td> <td>obs://mgc-test/data/migration-dli-spark-</td> <td>🗑</td> </tr> <tr> <td>mgc.mc2dli.data.migration.dli.spark.jars</td> <td>["obs://mgc-test/data/datasource.jar","oi"]</td> <td>🗑</td> </tr> </table> <p>添加参数</p> </div> <ul style="list-style-type: none"> - spark.dli.metaAccess.enable：填写true。 - spark.dli.job.agency.name：填写在完成准备工作时配置的DLI委托权限名称。 - mgc.mc2dli.data.migration.dli.file.path：填写在完成准备工作时存放migration-dli-spark-1.0.0.jar包的OBS路径。例如： obs://mgc-test/data/migration-dli-spark-1.0.0.jar - mgc.mc2dli.data.migration.dli.spark.jars：填写在完成准备工作时存放fastjson-1.2.54.jar和datasource.jar包的OBS路径，以数组形式传入，路径带双引号，以英文逗号隔开。例如：["obs://mgc-test/data/datasource.jar","obs://mgc-test/data/fastjson-1.2.54.jar"] 如果是专线迁移，需要填写以下8个参数： <div data-bbox="858 1236 1372 1601"> <p>自定义参数 参数详情请参考文档中“表3 conf参数说明”部分。 查看文档</p> <table border="1"> <tr> <td>spark.dli.metaAccess.enable</td> <td>true</td> <td>🗑</td> </tr> <tr> <td>spark.dli.job.agency.name</td> <td></td> <td>🗑</td> </tr> <tr> <td>mgc.mc2dli.data.migration.dli.file.path</td> <td>obs://mgc-test/data/migration-dli-spark-</td> <td>🗑</td> </tr> <tr> <td>mgc.mc2dli.data.migration.dli.spark.jars</td> <td>["obs://mgc-test/data/datasource.jar","oi"]</td> <td>🗑</td> </tr> <tr> <td>spark.sql.catalog.mc_catalog.tableWrite</td> <td>tunnel</td> <td>🗑</td> </tr> <tr> <td>spark.sql.catalog.mc_catalog.tableReac</td> <td>tunnel</td> <td>🗑</td> </tr> <tr> <td>spark.hadoop.odps.end.point</td> <td>http://service.ci. maxcompute</td> <td>🗑</td> </tr> <tr> <td>spark.hadoop.odps.tunnel.end.point</td> <td>http://dt.ci. j.maxcompute.allyt</td> <td>🗑</td> </tr> </table> <p>添加参数</p> </div> <ul style="list-style-type: none"> - spark.dli.metaAccess.enable：填写true。 - spark.dli.job.agency.name：填写在完成准备工作时配置的DLI委托权限名称。 - mgc.mc2dli.data.migration.dli.file.path：填写在完成准备工作时存放migration-dli-spark-1.0.0.jar包的OBS路径。例如： obs://mgc-test/data/migration-dli-spark-1.0.0.jar - mgc.mc2dli.data.migration.dli.spark.jars：填写在完成准备工作时存放 	spark.dli.metaAccess.enable	true	🗑	spark.dli.job.agency.name		🗑	mgc.mc2dli.data.migration.dli.file.path	obs://mgc-test/data/migration-dli-spark-	🗑	mgc.mc2dli.data.migration.dli.spark.jars	["obs://mgc-test/data/datasource.jar","oi"]	🗑	spark.dli.metaAccess.enable	true	🗑	spark.dli.job.agency.name		🗑	mgc.mc2dli.data.migration.dli.file.path	obs://mgc-test/data/migration-dli-spark-	🗑	mgc.mc2dli.data.migration.dli.spark.jars	["obs://mgc-test/data/datasource.jar","oi"]	🗑	spark.sql.catalog.mc_catalog.tableWrite	tunnel	🗑	spark.sql.catalog.mc_catalog.tableReac	tunnel	🗑	spark.hadoop.odps.end.point	http://service.ci. maxcompute	🗑	spark.hadoop.odps.tunnel.end.point	http://dt.ci. j.maxcompute.allyt	🗑
spark.dli.metaAccess.enable	true	🗑																																				
spark.dli.job.agency.name		🗑																																				
mgc.mc2dli.data.migration.dli.file.path	obs://mgc-test/data/migration-dli-spark-	🗑																																				
mgc.mc2dli.data.migration.dli.spark.jars	["obs://mgc-test/data/datasource.jar","oi"]	🗑																																				
spark.dli.metaAccess.enable	true	🗑																																				
spark.dli.job.agency.name		🗑																																				
mgc.mc2dli.data.migration.dli.file.path	obs://mgc-test/data/migration-dli-spark-	🗑																																				
mgc.mc2dli.data.migration.dli.spark.jars	["obs://mgc-test/data/datasource.jar","oi"]	🗑																																				
spark.sql.catalog.mc_catalog.tableWrite	tunnel	🗑																																				
spark.sql.catalog.mc_catalog.tableReac	tunnel	🗑																																				
spark.hadoop.odps.end.point	http://service.ci. maxcompute	🗑																																				
spark.hadoop.odps.tunnel.end.point	http://dt.ci. j.maxcompute.allyt	🗑																																				

区域	参数	配置说明
		<p>fastjson-1.2.54.jar和datasource.jar包的OBS路径，以数组形式传入，路径带双引号，以英文逗号隔开。例如：["obs://mgc-test/data/datasource.jar","obs://mgc-test/data/fastjson-1.2.54.jar"]</p> <ul style="list-style-type: none"> - spark.sql.catalog.mc_catalog.tableWriteProvider: 填写tunnel。 - spark.sql.catalog.mc_catalog.tableReadProvider: 填写tunnel。 - spark.hadoop.odps.end.point: 填写源端MaxCompute服务所在地域的VPC网络Endpoint信息。各地域VPC网络Endpoint信息请参考Endpoint对照表。例如：源端所在地域为“香港”，则填写“http://service.cn-hongkong.maxcompute.aliyun-inc.com/api” - spark.hadoop.odps.tunnel.end.point: 填写源端MaxCompute服务所在地域的VPC网络Tunnel Endpoint信息。各地域VPC网络Tunnel Endpoint信息请参考Endpoint对照表。例如：源端所在地域为“香港”，则填写“http://dt.cn-hongkong.maxcompute.aliyun-inc.com”
迁移选项	大表规则	设置大表条件，当符合大表的条件时，将自动按分区拆分迁移子任务。建议保持默认，也可根据实际需求修改。
	小表规则	设置小表条件，当符合小表的条件时，自动将多个小表合并到一个迁移子任务中，提高迁移效率。建议保持默认，也可根据实际需求修改。
	并发连接数	设置迁移并发数。默认为3，1≤取值范围≤10。
	单文件SQL数	每批提交的文件中所包含的SQL条数。默认为3，1≤推荐取值范围≤50。

步骤6 配置完成后，选择任务的运行方式。

须知

- 创建的迁移任务可重复运行，每运行一次，即会生成一条任务实例。
- 单击任务名称，可以修改任务配置。
- **立即运行**：如果需要立即运行迁移任务，可以勾选“**保存后立即运行**”，然后单击“**保存**”按钮，会在**任务管理**列表中生成一条全量数据迁移任务并自动开始执行迁移。



- **手动运行**：如果需要后续手动启动迁移任务，则只需单击“保存”按钮，在任务管理列表中生成一条全量数量迁移任务。在需要启动迁移时，单击操作列的“运行”，手动启动迁移。



步骤7 迁移任务运行后，单击操作列的“查看实例”，进入任务进度页签，可以查看该任务产生的运行实例以及迁移进度。



单击状态列的“重新运行”，可以多次运行迁移实例。



单击进度统计列的“查看”，进入进度详情页面。可以查看迁移的数据表清单、状态并导出任务结果。



步骤8 （可选）数据迁移完成后，可以使用MgC提供的大数据校验功能，对源端和目的端数据进行一致性校验。详细的操作步骤请参考[MaxCompute迁移至DLI数据校验](#)。

----结束

2.11.5.4 增量数据迁移

将上一次迁移后，在源端数据库中发生变化（新增、修改或删除）的数据同步至目的端华为云数据湖探索（DLI）。

前提条件

- 已完成所有准备工作。
- 已创建源端连接。
- 已创建目的端连接。
- 已最少完成一次全量数据迁移。
- 开始迁移前，需要联系技术支持人员，将您添加到允许jar程序访问DLI元数据功能的白名单中。

操作步骤

步骤1 登录迁移中心控制台。

步骤2 在左侧导航栏选择“实施>大数据数据迁移”，进入大数据数据迁移页面。在页面左上角的当前项目下拉列表中选择在完成准备工作时所创建的迁移项目。

步骤3 单击页面右上角的“创建迁移任务”按钮，右侧弹出创建迁移任务窗口。



步骤4 源端组件选择“MaxCompute”，目的端组件选择“数据湖探索 DLI”，任务类型选择“增量数据迁移”，单击“下一步”，进入详细配置页面。



步骤5 根据表2-69，配置增量数据迁移任务参数。

表 2-69 增量数据迁移参数配置说明

区域	参数	配置说明
基本信息	任务名称	默认创建名称为“MaxCompute到DLI增量数据迁移-4位随机数（字母和数字）”，也可以自定义名称。
	迁移工具	选择在完成准备工作时与迁移中心连接成功的Edge。
源端配置	源端连接	选择创建的源端连接。
	增量区间	选择T-N区间作为增量数据的迁移范围，即任务开始执行时间的前 24 x N 小时内的增量数据。例如：选择T-1，2024/06/06 14:50执行任务，则表示迁移从 2024/06/05 14:50 到 2024/06/06 14:50（现在）的增量数据。 如果勾选“指定日期”，将只迁移指定日期内的增量数据。
	分区筛选方式	根据数据的分区属性来选择和检索数据。默认为 更新时间 。 更新时间：筛选出最近或在特定时间段内有更新的数据。 创建时间：筛选出在特定时间窗口内创建的数据。
	MaxCompute参数（可选）	非必填参数，一般为空。如有需要，可参考 MaxCompute参数文档 进行配置。
数据范围	按库筛选	<p>即按数据库进行筛选，在数据库参数框中输入需要增量迁移的数据库名称。单击“添加数据库”，可以新增数据库，最多支持同时添加10个数据库。</p>  <p>如果数据库中有不需要迁移的表，可以下载CSV格式的“排除表”模板，参考b.~e.在模板中添加不需要迁移的表信息，然后上传至MgC。</p>

区域	参数	配置说明
	按表筛选	<ol style="list-style-type: none"> 1. 下载CSV格式的模板。 2. 下载后用记事本打开该CSV模板文件。 注意 请不要使用Excel对CSV模板文件进行编辑，通过Excel编辑并保存的模板文件无法被MgC识别。 3. 保留CSV文件中的第一行不变，从第二行开始，每行填写一个表的迁移信息，格式为“{MC项目名},{表名}”，其中MC项目名代表MaxCompute项目名称，表名代表具体的表名称。 须知 <ul style="list-style-type: none"> - 每行的“MC项目名”和“表名”之间使用英文逗号隔开，不要使用空格或其他分隔符。 - 每添加完一个表的信息后，使用回车键换行。 4. 完成所有表信息的添加后，保存CSV文件的更改。 5. 将编辑并保存好的CSV文件上传至MgC。
目的端配置	目的端连接	<p>选择创建的目的端DLI 通用队列连接。</p> <p>注意 切勿选择DLI SQL队列。</p>

区域	参数	配置说明																																				
	自定义参数（可选）	<p>根据实际需求进行配置，支持的自定义参数详情请参考conf参数说明和自定义参数说明。</p> <ul style="list-style-type: none"> 如果是公网迁移，需要填写以下4个参数： <div data-bbox="858 421 1372 631" style="border: 1px solid #ccc; padding: 5px; margin: 5px 0;"> <p>自定义参数 参数详情请参考文档中“表3 conf参数说明”部分。 查看文档</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 60%;">spark.dli.metaAccess.enable</td> <td>true</td> <td style="text-align: right;">✕</td> </tr> <tr> <td>spark.dli.job.agency.name</td> <td></td> <td style="text-align: right;">✕</td> </tr> <tr> <td>mgc.mc2dli.data.migration.dli.file.path</td> <td>obs://mgc-test/data/migration-dli-spark-</td> <td style="text-align: right;">✕</td> </tr> <tr> <td>mgc.mc2dli.data.migration.dli.spark.jars</td> <td>["obs://mgc-test/data/datasource.jar","oi"]</td> <td style="text-align: right;">✕</td> </tr> </table> <p style="text-align: center; color: #00a0e3; font-size: small;">添加参数</p> </div> <ul style="list-style-type: none"> - spark.dli.metaAccess.enable：填写true。 - spark.dli.job.agency.name：填写在完成准备工作时配置的DLI委托权限名称。 - mgc.mc2dli.data.migration.dli.file.path：填写在完成准备工作时存放migration-dli-spark-1.0.0.jar包的OBS路径。例如： obs://mgc-test/data/migration-dli-spark-1.0.0.jar - mgc.mc2dli.data.migration.dli.spark.jars：填写在完成准备工作时存放fastjson-1.2.54.jar和datasource.jar包的OBS路径，以数组形式传入，路径带双引号，以英文逗号隔开。例如：["obs://mgc-test/data/datasource.jar","obs://mgc-test/data/fastjson-1.2.54.jar"] 如果是专线迁移，需要填写以下8个参数： <div data-bbox="858 1236 1372 1601" style="border: 1px solid #ccc; padding: 5px; margin: 5px 0;"> <p>自定义参数 参数详情请参考文档中“表3 conf参数说明”部分。 查看文档</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 60%;">spark.dli.metaAccess.enable</td> <td>true</td> <td style="text-align: right;">✕</td> </tr> <tr> <td>spark.dli.job.agency.name</td> <td></td> <td style="text-align: right;">✕</td> </tr> <tr> <td>mgc.mc2dli.data.migration.dli.file.path</td> <td>obs://mgc-test/data/migration-dli-spark-</td> <td style="text-align: right;">✕</td> </tr> <tr> <td>mgc.mc2dli.data.migration.dli.spark.jars</td> <td>["obs://mgc-test/data/datasource.jar","oi"]</td> <td style="text-align: right;">✕</td> </tr> <tr> <td>spark.sql.catalog.mc_catalog.tableWrite</td> <td>tunnel</td> <td style="text-align: right;">✕</td> </tr> <tr> <td>spark.sql.catalog.mc_catalog.tableReac</td> <td>tunnel</td> <td style="text-align: right;">✕</td> </tr> <tr> <td>spark.hadoop.odps.end.point</td> <td>http://service.ci. maxcompute</td> <td style="text-align: right;">✕</td> </tr> <tr> <td>spark.hadoop.odps.tunnel.end.point</td> <td>http://dt.ci. j.maxcompute.allyt</td> <td style="text-align: right;">✕</td> </tr> </table> <p style="text-align: center; color: #00a0e3; font-size: small;">添加参数</p> </div> <ul style="list-style-type: none"> - spark.dli.metaAccess.enable：填写true。 - spark.dli.job.agency.name：填写在完成准备工作时配置的DLI委托权限名称。 - mgc.mc2dli.data.migration.dli.file.path：填写在完成准备工作时存放migration-dli-spark-1.0.0.jar包的OBS路径。例如： obs://mgc-test/data/migration-dli-spark-1.0.0.jar - mgc.mc2dli.data.migration.dli.spark.jars：填写在完成准备工作时存放 	spark.dli.metaAccess.enable	true	✕	spark.dli.job.agency.name		✕	mgc.mc2dli.data.migration.dli.file.path	obs://mgc-test/data/migration-dli-spark-	✕	mgc.mc2dli.data.migration.dli.spark.jars	["obs://mgc-test/data/datasource.jar","oi"]	✕	spark.dli.metaAccess.enable	true	✕	spark.dli.job.agency.name		✕	mgc.mc2dli.data.migration.dli.file.path	obs://mgc-test/data/migration-dli-spark-	✕	mgc.mc2dli.data.migration.dli.spark.jars	["obs://mgc-test/data/datasource.jar","oi"]	✕	spark.sql.catalog.mc_catalog.tableWrite	tunnel	✕	spark.sql.catalog.mc_catalog.tableReac	tunnel	✕	spark.hadoop.odps.end.point	http://service.ci. maxcompute	✕	spark.hadoop.odps.tunnel.end.point	http://dt.ci. j.maxcompute.allyt	✕
spark.dli.metaAccess.enable	true	✕																																				
spark.dli.job.agency.name		✕																																				
mgc.mc2dli.data.migration.dli.file.path	obs://mgc-test/data/migration-dli-spark-	✕																																				
mgc.mc2dli.data.migration.dli.spark.jars	["obs://mgc-test/data/datasource.jar","oi"]	✕																																				
spark.dli.metaAccess.enable	true	✕																																				
spark.dli.job.agency.name		✕																																				
mgc.mc2dli.data.migration.dli.file.path	obs://mgc-test/data/migration-dli-spark-	✕																																				
mgc.mc2dli.data.migration.dli.spark.jars	["obs://mgc-test/data/datasource.jar","oi"]	✕																																				
spark.sql.catalog.mc_catalog.tableWrite	tunnel	✕																																				
spark.sql.catalog.mc_catalog.tableReac	tunnel	✕																																				
spark.hadoop.odps.end.point	http://service.ci. maxcompute	✕																																				
spark.hadoop.odps.tunnel.end.point	http://dt.ci. j.maxcompute.allyt	✕																																				

区域	参数	配置说明
		<p>fastjson-1.2.54.jar和datasource.jar包的OBS路径，以数组形式传入，路径带双引号，以英文逗号隔开。例如：["obs://mgc-test/data/datasource.jar","obs://mgc-test/data/fastjson-1.2.54.jar"]</p> <ul style="list-style-type: none"> - spark.sql.catalog.mc_catalog.tableWriteProvider: 填写tunnel。 - spark.sql.catalog.mc_catalog.tableReadProvider: 填写tunnel。 - spark.hadoop.odps.end.point: 填写源端MaxCompute服务所在地域的VPC网络Endpoint信息。各地域VPC网络Endpoint信息请参考Endpoint对照表。例如：源端所在地域为“香港”，则填写“http://service.cn-hongkong.maxcompute.aliyun-inc.com/api” - spark.hadoop.odps.tunnel.end.point: 填写源端MaxCompute服务所在地域的VPC网络Tunnel Endpoint信息。各地域VPC网络Tunnel Endpoint信息请参考Endpoint对照表。例如：源端所在地域为“香港”，则填写“http://dt.cn-hongkong.maxcompute.aliyun-inc.com”
迁移选项	大表规则	设置大表条件，当符合大表的条件时，将自动按分区拆分迁移子任务。建议保持默认，也可根据实际需求修改。
	小表规则	设置小表条件，当符合小表的条件时，自动将多个小表合并到一个迁移子任务中，提高迁移效率。建议保持默认，也可根据实际需求修改。
	并发连接数	设置迁移并发数。默认为3，1≤取值范围≤10。
	单文件SQL数	每批提交的文件中所包含的SQL条数。默认为3，1≤推荐取值范围≤50。

步骤6 配置完成后，选择任务的运行方式。

须知

- 创建的迁移任务可重复运行，每运行一次，即会生成一条任务实例。
- 单击任务名称，可以修改任务配置。
- **立即运行**：如果需要立即运行迁移任务，可以勾选“**保存后立即运行**”，然后单击“**保存**”按钮，会在**任务管理**列表中生成一条增量数据迁移任务并自动开始执行迁移。



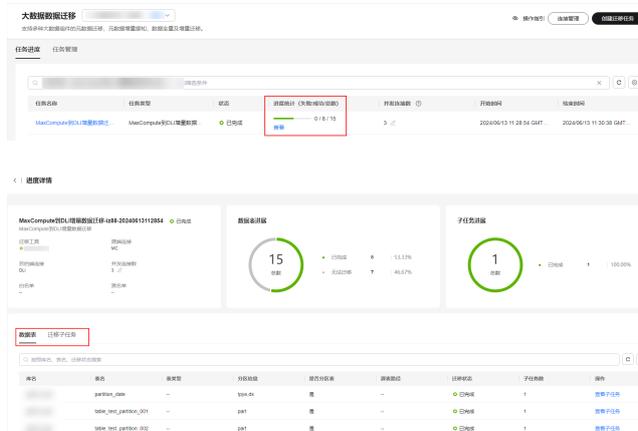
- **手动运行**：如果需要后续手动启动迁移任务，则只需单击“保存”按钮，在任务管理列表中生成一条增量数据迁移任务。在需要启动迁移时，单击操作列的“运行”，手动启动迁移。



步骤7 迁移任务运行后，单击操作列的“查看实例”，进入任务进度页签，可以查看该任务的迁移进度以及产生的所有运行实例。



单击进度统计列的“查看”，进入进度详情页面。可以查看迁移的数据表清单、状态并导出任务结果。



步骤8 （可选）数据迁移完成后，可以使用MgC提供的大数据校验功能，对源端和目的端数据进行一致性校验。详细的操作步骤请参考[MaxCompute迁移至DLI数据校验](#)。

----结束

2.11.6 查看并导出执行结果

每运行一次迁移任务，即会生成一条任务实例。在任务实例的进度详情页面，可以查看并导出任务的执行结果。

须知

导出的结果文件可以在生成的15天内下载，15天后自动删除。

前提条件

已创建迁移任务并运行。

操作步骤

步骤1 在任务进度列表，单击进度统计列的“查看”，进入**进度详情**页面。



步骤2 单击列表上方的“导出”按钮，系统开始导出任务执行结果。通过列表上方的搜索框可以对任务执行结果进行过滤，如有过滤，则只导出过滤后的数据。

图 2-36 导出执行结果



步骤3 单击“导出管理”按钮，右侧弹出**导出管理**窗口。

步骤4 当导出进度为导出成功后，单击操作列的“下载”，可以将导出结果下载到本地。

图 2-37 下载导出结果



---结束

2.11.7 自定义参数说明

在目的端配置中，支持配置的自定义参数及说明请参见[表2-70](#)和[conf参数说明](#)。

表 2-70 支持的自定义参数说明

参数名称	取值范围	说明	是否必填
spark.dli.metaAccess.enable	-	填写true。	是
spark.dli.job.agency.name	-	填写 DLI委托权限名称 。	
mgc.mc2dli.data.migration.dli.file.path	-	填写 存放migration-dli-spark-1.0.0.jar包的OBS路径 。例如： obs://mgc-test/data/migration-dli-spark-1.0.0.jar	
mgc.mc2dli.data.migration.dli.spark.jars	-	填写 存放fastjson-1.2.54.jar和datasource.jar包的OBS路径 ，以数组形式传入，路径带双引号，以英文逗号隔开。例如：["obs://mgc-test/data/datasource.jar","obs://mgc-test/data/fastjson-1.2.54.jar"]	
spark.sql.catalog.mc_catalog.tableWriteProvider	-	填写tunnel	专线迁移场景为必填
spark.sql.catalog.mc_catalog.tableReadProvider	-	填写tunnel	

参数名称	取值范围	说明	是否必填
spark.hadoop.odps.endpoint		填写源端MaxCompute服务所在地域的 VPC网络Endpoint 信息。各地域 VPC网络Endpoint 信息请参考 Endpoint对照表 。例如：源端所在地域为“香港”，则填写“http://service.cn-hongkong.maxcompute.aliyun-inc.com/api”	
spark.hadoop.odps.tunnel.endpoint		填写源端MaxCompute服务所在地域的 VPC网络Tunnel Endpoin 信息。各地域 VPC网络Tunnel Endpoin 信息请参考 Endpoint对照表 。例如：源端所在地域为“香港”，则填写“http://dt.cn-hongkong.maxcompute.aliyun-inc.com”	
spark.hadoop.odps.tunnel.quota.name	-	填写访问MaxCompute使用的Quota名称。	否
sc_type	A、B、C	<p>计算资源类型，目前指定参数A, B, C。如果不指定，则按最小类型创建。</p> <ul style="list-style-type: none"> ● A: (8核32G内存, driverCores: 2, executorCores: 1, driverMemory: 7G, executorMemory: 4G, numExecutor: 6) ● B: (16核64G内存, driverCores: 2, executorCores: 2, driverMemory: 7G, executorMemory: 8G, numExecutor: 7) ● C: (32核128G内存, driverCores: 4, executorCores: 2, driverMemory: 15G, executorMemory: 8G, numExecutor: 14) 	否
executorCores	1- 4	Spark应用每个Executor的CPU核数。该配置项会替换“sc_type”中对应的默认参数。	
numExecutors	1-100	Spark应用Executor的个数。该配置项会替换“sc_type”中对应的默认参数。	

参数名称	取值范围	说明	是否必填
executorMemory	1-1 6 G	Spark应用的Executor内存，参数配置例如：2G, 2048M。该配置项会替换“sc_type”中对应的默认参数，使用时必须带单位，否则会启动失败。	
driverCores	1-4	Spark应用Driver的CPU核数。该配置项会替换“sc_type”中对应的默认参数。	
driverMemory	1-1 6 G	Spark应用的Driver内存，参数配置例如：2G, 2048M。该配置项会替换“sc_type”中对应的默认参数，使用时必须带单位，否则会启动失败。	

2.12 大数据校验

2.12.1 大数据校验概述

迁移中心 MgC为多种大数据计算/存储引擎，如Hive、Hbase、Doris和MaxCompute等，提供了一致性校验方法。通过一致性校验，您可以更加放心地进行大数据迁移操作，确保数据的准确性和可靠性，从而避免了数据丢失或不一致的风险。

校验方式说明

- **全量校验：**全量对比所有数据，适用于需要对比历史数据的场景。
- **日级增量校验：**依靠元数据中的创建或更新时间，获取数据变化（增量）分区，再进行校验。支持单日或连续多日增量数据的统计，是最常用的校验方式。
- **小时级增量校验：**依靠元数据中的创建或更新时间，获取数据变化（增量）分区，再进行校验。在24小时内自动进行多次校验，可以持续关注数据一致性变化情况。第二日0点自动停止任务。
- **指定日期校验：**仅适用于分区为年月日格式的分表，支持周期性指定一个或多个连续日期分区进行校验。分区格式不为年月日的表不会进行校验。
- **指定时间校验：**用户可以选择一个时间段，校验该时间段内的数据。只能选择当前时间之前的时间段进行校验。

支持的源端、目的端组件

源端组件	目的端组件
<ul style="list-style-type: none"> • Hive • Hbase • Doris • MaxCompute • ClickHouse • Delta Lake • Hudi 	<ul style="list-style-type: none"> • Hive • DLI • MRS(Doris) • MRS(Hbase) • MRS(ClickHouse) • CloudTable(ClickHouse) • CloudTable(HBase) • Delta • Hudi

各组件支持的校验方式

组件	支持的校验方式
Hive	<ul style="list-style-type: none"> • 全量校验
DLI	<ul style="list-style-type: none"> • 日级增量校验 • 小时级增量校验 • 指定日期校验
MaxCompute	<ul style="list-style-type: none"> • 全量校验 • 日级增量校验 • 小时级增量校验 • 指定日期校验 • 抽样分区校验
Doris	<ul style="list-style-type: none"> • 全量校验 • 日级增量校验 • 小时级增量校验
Hbase	<ul style="list-style-type: none"> • 全量校验 • 指定时间校验
ClickHouse	全量校验
阿里云云数据库 ClickHouse	全量校验
CloudTable(HBase)	<ul style="list-style-type: none"> • 全量校验 • 指定时间校验
CloudTable(ClickHouse)	全量校验

组件	支持的校验方式
Delta	<ul style="list-style-type: none"> ● 全量校验 ● 日级增量校验 ● 小时级增量校验 ● 指定日期校验
Hudi	<ul style="list-style-type: none"> ● 全量校验 ● 日级增量校验 ● 小时级增量校验 ● 指定日期校验

2.12.2 准备工作

在使用大数据校验功能前，请务必先完成以下必要的准备工作。

创建项目

通过创建独立的项目，可以将不同的操作场景进行区分和隔离，从而实现更好的组织管理和资源分配。创建方法请参考[项目管理](#)。

准备迁移工具 Edge

在启用大数据校验功能前，需要先在源端环境中安装用于配合大数据校验的迁移工具 Edge。大数据校验场景仅支持Linux版本的Edge，安装方法请参考[下载并安装Linux版本Edge](#)。安装成功后，将Edge与迁移中心进行连接，连接方法请参考[连接迁移中心](#)。

注意

迁移工具Edge不具备自动重启功能，严禁在任务执行期间重启Edge，以避免任务失败。

选择并启用迁移工具

- 步骤1** 迁移工具Edge与迁移中心连接成功后，登录[迁移中心](#)管理控制台。
- 步骤2** 在左侧导航栏选择“**实施>大数据校验**”，进入**大数据校验**页面。在页面左上角的**当前项目**下拉列表中选择**创建的项目**。
- 步骤3** 单击“**选择并启用迁移工具**”按钮，弹出**选择迁移工具**窗口。
- 步骤4** 在下拉列表中，选择与迁移中心连接成功的Edge。
 - 如迁移工具状态为“**在线**”、“**已启用**”，代表该工具已被用于对数相关操作，单击“**确定**”按钮，继续使用该工具进行后续操作。
 - 如迁移工具状态为“**在线**”、“**未启用**”，代表该工具尚未被用于对数，单击“**启用**”按钮，启用该工具进行后续操作。

- 如迁移工具状态为“**离线**”、“**已删除**”，代表该工具已被删除，无法进行对数操作；如果该工具在删除前进行过对数操作，可以单击“**查看**”按钮，查看已有数据。
- 如迁移工具状态为“**离线**”、“**未启用**”，代表该工具无法进行对数相关操作，单击“**查看**”按钮，界面上的所有操作按钮会被禁用。

 **注意**

在进行大数据校验前，请确保迁移工具Edge处于在线并启用状态。

----结束

2.12.3 创建源端连接

2.12.3.1 创建源端云服务连接

对大数据云服务数据进行一致性校验时，需要分别与源端和目的端云服务建立连接，用于数据查询和计算。

目前支持创建源端连接的大数据云服务包括：

- **MaxCompute**
- **阿里云云数据库ClickHouse**

操作步骤

步骤1 登录[迁移中心](#)管理控制台。

步骤2 在左侧导航栏选择“**实施>大数据校验**”，进入**大数据校验**页面。在页面左上角的**当前项目**下拉列表中选择**迁移项目**。

步骤3 在**功能入口**区域，单击“**连接管理**”，进入**连接管理**页面。

步骤4 单击页面右上角的“**新增连接**”，右侧弹出**新增连接**窗口。

步骤5 根据实际情况选择大数据云服务，单击“**下一步**”，进入**详细配置**页面。

步骤6 根据对应的参数配置说明，配置云服务参数。

- [MaxCompute连接参数配置说明](#)
- [阿里云云数据库ClickHouse连接参数配置说明](#)

表 2-71 MaxCompute 连接参数配置说明

参数	配置说明
所属源	选择源端。
连接名称	默认创建名称为“MaxCompute-4位随机数（字母和数字）”，也可以自定义名称。
阿里云凭证	选择源端阿里云账号的凭证。凭证的添加方式请参考 添加资源凭证 ，类型选择“大数据-MaxCompute”。

参数	配置说明
项目名称	输入需要采集的MaxCompute项目名称。项目名称可在MaxCompute侧控制台获取。
Endpoint	输入MaxCompute服务所在地域的Endpoint信息。各地域Endpoint信息请参考 Endpoint对照表 。

表 2-72 阿里云云数据库 ClickHouse 连接参数配置说明

参数	配置说明
所属源	选择源端。
连接名称	默认创建名称为“阿里云云数据库ClickHouse-4位随机数（字母和数字）”，也可以自定义名称。
ClickHouse凭证	选择在Edge添加的源端阿里云云数据库ClickHouse凭证。凭证的添加方法请参考 添加资源凭证 中的大数据-ClickHouse。
数据库链接	填写连接源端ClickHouse集群的外网地址。可以在 集群信息 页面，查看外网地址。

步骤7 参数配置完成后，单击“**测试连接**”，系统会校验配置信息并与云服务建立连接，测试通过代表与云服务连接成功。

步骤8 测试连接通过后，单击“**确认**”按钮，创建云服务连接完成。

步骤9 在**连接管理**页面，可以查看创建的连接列表以及基本信息。单击操作列的“**修改连接**”，可以修改连接参数配置。

----结束

2.12.3.2 创建源端组件连接

对大数据组件数据进行一致性校验时，需要与大数据组件建立连接，用于数据查询。

目前支持创建源端连接的大数据组件包括：

- Doris
- HBase
- ClickHouse
- Hive Metastore
- Delta Lake（有元数据）
- Delta Lake（无元数据）
- Hudi（有元数据）
- Hudi（无元数据）

操作步骤

- 步骤1** 登录[迁移中心](#)管理控制台。
- 步骤2** 在左侧导航栏选择“[实施>大数据校验](#)”，进入[大数据校验](#)页面。在页面左上角的[当前项目](#)下拉列表中选择[迁移项目](#)。
- 步骤3** 在功能入口区域，单击“[连接管理](#)”，进入[连接管理](#)页面。
- 步骤4** 单击页面右上角的“[新增连接](#)”，右侧弹出[新增连接](#)窗口。
- 步骤5** 根据实际情况选择大数据组件，单击“[下一步](#)”，进入[详细配置](#)页面。
- 步骤6** 根据对应的参数配置说明，配置执行机参数。
- [Doris连接参数配置说明](#)
 - [HBase连接参数配置说明](#)
 - [ClickHouse组件参数配置说明](#)
 - [Hive Metastore连接参数配置说明](#)
 - [Delta Lake（有元数据）连接参数配置说明](#)
 - [Delta Lake（无元数据）连接参数配置说明](#)
 - [Hudi（有元数据）连接参数配置说明](#)
 - [Hudi（无元数据）连接参数配置说明](#)

表 2-73 Doris 连接参数配置说明

参数	配置说明
所属源	选择源端。
连接名称	默认创建名称为“Doris-4位随机数（字母和数字）”，也可以自定义名称。
Doris凭证	选择在Edge添加的源端Doris凭证。凭证的添加方式请参考 添加资源凭证 ，类型选择“大数据-Doris”。
数据库地址	填写连接到源端Doris集群部署的服务器IP地址。
数据库端口	填写连接到源端Doris集群部署的服务器的端口。默认为3306。
数据库名	填写源端Doris数据库名称。

表 2-74 HBase 连接参数配置说明

参数	配置说明
所属源	选择源端。
连接名称	默认创建名称为“HBase-4位随机数（字母和数字）”，也可以自定义名称。

参数	配置说明
HBase凭证	选择在Edge添加的源端HBase凭证。凭证的添加方式请参考 添加资源凭证 ，类型选择“大数据-HBase”。
安全集群	选择是否部署在安全集群下。
Zookeeper节点IP	填写连接到源端Zookeeper的IP地址。通常是部署Zookeeper的服务器公网IP地址或内网IP地址。
Zookeeppeer连接端口	填写连接到源端Zookeeppeer的端口。
HBase 版本	选择源端HBase的版本。

表 2-75 ClickHouse 连接参数配置说明

参数	配置说明
所属源	选择源端。
连接名称	默认创建名称为“ClickHouse-4位随机数（字母和数字）”，也可以自定义名称。
ClickHouse凭证（可选）	选择在Edge添加的源端ClickHouse凭证。凭证的添加方法请参考 添加资源凭证 中的大数据-ClickHouse。
安全集群	选择是否部署在安全集群下。
ClickHouse Server IP	填写连接到源端ClickHouse服务器的IP地址。通常是运行ClickHouse的主机IP地址。
HTTP端口	非安全集群下，与源端ClickHouse 服务器通信的 HTTP 端口。默认为8123。
HTTP SSL/TLS端口	安全集群下，与源端ClickHouse 服务器通信的HTTPS 端口。

表 2-76 Hive Metastore 连接参数配置说明

参数	配置说明
所属源	选择源端。
连接名称	默认创建名称为“Hive-Metastore-4位随机数（字母和数字）”，也可以自定义名称。
启用安全连接	选择是否启用安全连接。 <ul style="list-style-type: none"> MetaStore在非安全集群下，可以不启用安全连接。 MetaStore在安全集群下，需要启用安全连接并选择凭证。凭证的获取以及添加方法请参考添加资源凭证中的大数据-Hive Metastore。

参数	配置说明
Hive版本	选择源端Hive的实际版本。 注意 如果源端Hive为2.1.1版本，此处需要选择1.x版本。
MetaStore节点IP	填写连接MetaStore的IP地址。
MetaStore Thrift服务端口	填写连接MetaStore的端口，默认为9083。
启用元数据仓库连接	<p>在增量数据校验中，如果通过Metastore查询，当分区数量超过30,000时，可能会因为将所有分区信息加载到内存中而导致内存溢出（OOM）。因此，使用MySQL作为元数据仓库可以有效地解决这一问题。</p> <ul style="list-style-type: none"> 不启用：系统通过MetaStore查询Hive表的分区信息。 启用：配置MySQL作为元数据仓库，系统将通过MySQL查询Hive表的分区信息。启用后，需要配置如下参数： <ul style="list-style-type: none"> 元仓类型：目前仅支持MySQL作为元数据仓库。 MySQL凭证：选择访问MySQL数据库的认证信息。认证信息需要添加到Edge然后同步到MgC，方法请参考添加资源凭证。 MySQL节点IP：输入MySQL数据库服务器的IP地址。 MySQL服务端口：输入MySQL数据库服务的端口号。 数据库名：输入存储Hive表元数据的数据库名称。 <p>说明 请确保填写的MySQL凭证、节点IP、服务端口和数据库名与Hive实际使用的MySQL数据库相匹配，否则会导致数据校验失败。</p>

表 2-77 Delta Lake（有元数据）连接参数配置说明

参数	配置说明
所属源	选择源端。
连接名称	默认创建名称为“Delta-Lake-有元数据-4位随机数（字母和数字）”，也可以自定义名称。
执行机凭证	选择执行机的登录凭证。凭证的添加方法请参考 添加资源凭证 中的大数据-执行机。
执行机IP	填写连接执行机的IP地址。
执行机连接端口	填写连接执行机的端口。
Spark客户端目录	填写Spark客户端的安装目录路径。
环境变量地址	填写环境变量文件（配置文件）的绝对路径，例如：/opt/bigdata/client/bigdata_env。

参数	配置说明
SQL文件地址	<p>填写用于存放执行作业时生成的SQL文件的文件夹地址。您需要拥有该文件夹的读写权限。</p> <p>须知 迁移完成后，您需要手动清理该路径下生成的文件夹，以释放存储空间。</p>

表 2-78 Delta Lake（无元数据）连接参数配置说明

参数	配置说明
所属源	选择源端。
连接名称	默认创建名称为“Delta-Lake-无元数据-4位随机数（字母和数字）”，也可以自定义名称。
执行机凭证	选择执行机的登录凭证。凭证的添加方法请参考 添加资源凭证 中的大数据-执行机。
执行机IP	填写连接执行机的IP地址。
执行机连接端口	填写连接执行机的端口。
Spark客户端目录	填写Spark客户端的安装目录路径。
环境变量地址	填写环境变量文件（配置文件）的绝对路径，例如： <code>/opt/bigdata/client/bigdata_env</code> 。
SQL文件地址	<p>填写用于存放执行作业时生成的SQL文件的文件夹地址。您需要拥有该文件夹的读写权限。</p> <p>须知 迁移完成后，您需要手动清理该路径下生成的文件夹，以释放存储空间。</p>

表 2-79 Hudi（有元数据）连接参数配置说明

参数	配置说明
所属源	选择源端。
连接名称	默认创建名称为“Hudi-有元数据-4位随机数（字母和数字）”，也可以自定义名称。
执行机凭证	选择执行机的登录凭证。凭证的添加方法请参考 添加资源凭证 中的大数据-执行机。
执行机IP	填写连接执行机的IP地址。
执行机连接端口	填写连接执行机的端口。
Spark客户端目录	填写Spark客户端的安装目录路径。

参数	配置说明
环境变量地址	填写环境变量文件（配置文件）的绝对路径，例如：/opt/bigdata/client/bigdata_env。
SQL文件地址	填写用于存放执行作业时生成的SQL文件的文件夹地址。您需要拥有该文件夹的读写权限。 须知 迁移完成后，您需要手动清理该路径下生成的文件夹，以释放存储空间。

表 2-80 Hudi（无元数据）连接参数配置说明

参数	配置说明
所属源	选择源端。
连接名称	默认创建名称为“Hudi-无元数据-4位随机数（字母和数字）”，也可以自定义名称。
执行机凭证	选择执行机的登录凭证。凭证的添加方法请参考 添加资源凭证 中的大数据-执行机。
执行机IP	填写连接执行机的IP地址。
执行机连接端口	填写连接执行机的端口。
Spark客户端目录	填写Spark客户端的安装目录路径。
环境变量地址	填写环境变量文件（配置文件）的绝对路径，例如：/opt/bigdata/client/bigdata_env。
SQL文件地址	填写用于存放执行作业时生成的SQL文件的文件夹地址。您需要拥有该文件夹的读写权限。 须知 迁移完成后，您需要手动清理该路径下生成的文件夹，以释放存储空间。

步骤7 参数配置完成后，单击“**测试连接**”，系统会校验配置信息并与组件建立连接，测试通过代表与组件连接成功。

步骤8 测试连接通过后，单击“**确认**”按钮，创建大数据组件连接完成。

步骤9 在**连接管理**页面，可以查看创建的连接列表以及基本信息。单击操作列的“**修改连接**”，可以修改连接参数配置。

----结束

2.12.4 创建表组并添加数据表

将源端数据表加入表组中，为后续对数作业做准备。

使用须知

- 每次最多支持导入10,000张表。
- 待导入的表必须属于同一个元数据来源。
- 导入数据表时，系统不会校验数据表加入的分组数量和对数标准。为保持对数的准确性，建议同一数据表加入的分组不超过3个，并避免加入对数标准相同的分组。
- 在创建用于MaxCompute迁移到DLI数据校验的表组时，建议先创建**count**、**sum**、**allsum**等基本校验标准的表组，在确保基本校验任务的校验标准已达标，并且源端和目的端数据量一致后，再创建**content**校验标准的表组进行详细的内容校验，以确保的目的端数据与源数据在数据量和内容上一致。

前提条件

- 已完成[准备工作](#)。
- 已[创建源端连接](#)。

创建表组

- 步骤1** 登录[迁移中心](#)管理控制台。
- 步骤2** 在左侧导航栏选择“[实施](#)>[大数据校验](#)”，进入[大数据校验](#)页面。在页面左上角的**当前项目**下拉列表中选择[迁移项目](#)。
- 步骤3** 在**功能入口**区域，单击“[表管理](#)”，进入表管理页面的**表组**页签。
- 步骤4** 单击“[新增表组](#)”，弹出[新增表组](#)窗口。
- 步骤5** 根据[表2-81](#)，配置参数。

表 2-81 新增表组参数说明

参数	说明
表组名	用户自定义
元数据连接	根据实际需求，选择 创建的源端连接 。 注意 表组中的所有表必须来源于同一个元数据来源。
校验标准	选择可以接受的数据差异最大值（即错误容忍度）。MgC提供了多种类型的对数标准可供选择，关于对数标准的详细说明，可以单击新增表组弹窗中的“ 查看更多 ”进行查看。
说明（可选）	您可以输入用于区分表组的说明。

- 步骤6** 参数配置完成后，单击“[确认](#)”按钮，创建表组完成。在表组列表，可以查看创建的表组信息。

表组创建完成后，即可[导入数据表并加入表组](#)。

----**结束**

导入数据表并加入表组

- 步骤1** 在**表管理**页面，选择**库表清单**页签。
- 步骤2** 单击列表上方的“**表管理 > 导入数据表**”按钮，弹出**导入数据表**窗口。
- 步骤3** 选择元数据连接和要加入的表组名。
- 步骤4** 单击“**模板下载**”，将导入模板下载到本地。打开导入模板，将需要导入的数据表对应参数填写到模板内。

须知

- 每个模板最多支持导入10,000张表。
- 同一表组内的表必须属于同一个元数据来源。
- 模板中单元格不允许存在公式且必须是文本格式，否则会导致解析失败。
- 元数据连接为**Delta（有元数据）**或**Hudi（有元数据）**连接时，模板中的**source_path**参数为必填项。
- 元数据连接为**Delta（无元数据）**或**Hudi（无元数据）**连接时，模板中的**source_path**和**target_path**参数为必填项。

- 步骤5** 模板参数填写完成并保存后，单击导入窗口的“**添加文件**”按钮，将填写后的模板文件上传。
- 步骤6** 上传成功后，单击“**确认**”按钮，系统开始导入。在库表清单列表中可以查看导入的数据表。

----结束

导出数据表

支持将Delta Lake和Hudi有**元数据**的数据表导出为CSV格式。无元数据的数据表不支持导出。

- 步骤1** 在**表管理**页面，选择**库表清单**页签。
- 步骤2** 单击列表上方的“**导出 > 导出数据表**”按钮，弹出**导出**窗口。



- 步骤3** 选择需要导出的数据表所在表组，单击“**确认**”。

步骤4 导出完成后，单击“**导出 > 导出管理**”按钮，右侧弹出**导出管理**窗口。

步骤5 单击操作列的“**下载**”，可以将导出的数据表下载到本地。

----结束

2.12.5 创建目的端连接

2.12.5.1 创建执行机连接

执行机是获取数据的任务执行单元，需要分别与源端和目的端的执行机建立连接，然后执行数据统计，统计结果将用于数据校验。

以下大数据组件需要通过创建执行机连接来进行数据校验：

- **Hive Metastore**
- **Delta Lake**（有元数据）
- **Delta Lake**（无元数据）
- **Hudi**（有元数据）
- **Hudi**（无元数据）

注意

- 需要确保源端和目的端执行机都安装了大数据组件的命令行客户端。
- 源端和目的端执行机均需要创建连接。

操作步骤

步骤1 登录**迁移中心**管理控制台。

步骤2 在左侧导航栏选择“**实施>大数据校验**”，进入**大数据校验**页面。在页面左上角的**当前项目**下拉列表中选择**迁移项目**。

步骤3 在功能入口区域，单击“**连接管理**”，进入**连接管理**页面。

步骤4 单击页面右上角的“**新增连接**”，右侧弹出**新增连接**窗口。

步骤5 根据实际情况选择执行机类型，单击“**下一步**”，根据**表2-82**，配置执行机参数。

表 2-82 执行机参数配置说明

参数	配置说明
所属源	根据实际情况选择。 须知 CDH执行机只支持源端，MRS执行机支持源端和目的端。
连接名称	默认创建名称为“执行机类型-4位随机数（字母和数字）”，也可以自定义名称。

参数	配置说明
执行机凭证	选择执行机的登录凭证。凭证的添加方法请参考 添加资源凭证 中的大数据-执行机。
执行机IP	填写连接执行机的IP地址。
执行机连接端口	填写连接执行机的端口，默认为22端口。
安装目录	填写MRS/CDH客户端的安装目录路径。即安装客户端时，“./install.sh”所填写的地址。
SQL文件地址	填写用于存放执行作业时生成的SQL文件的文件夹地址。您需要拥有该文件夹的读写权限。 须知 校验完成后，您需要手动清理该路径下生成的文件夹，以释放存储空间。

步骤6 参数配置完成后，单击“[测试连接](#)”，系统会校验配置信息并与执行机建立连接，测试通过代表与执行机连接成功。

步骤7 测试连接通过后，单击“[确认](#)”按钮，创建执行机连接完成。

步骤8 在[连接管理](#)页面，可以查看创建的执行机连接列表以及基本信息。单击操作列的“[修改连接](#)”，可以修改连接参数配置。

----结束

2.12.5.2 创建目的端云服务连接

对大数据云服务数据进行一致性校验时，需要分别与源端和目的端云服务建立连接，用于数据查询和计算。

目前支持创建目的端连接的大数据云服务包括：

- [数据湖探索 \(DLI\)](#)
- [CloudTable\(ClickHouse\)](#)
- [CloudTable\(HBase\)](#)

操作步骤

步骤1 登录[迁移中心](#)管理控制台。

步骤2 在左侧导航栏选择“[实施>大数据校验](#)”，进入[大数据校验](#)页面。在页面左上角的当前项目下拉列表中选择[迁移项目](#)。

步骤3 在功能入口区域，单击“[连接管理](#)”，进入[连接管理](#)页面。

步骤4 单击页面右上角的“[新增连接](#)”，右侧弹出[新增连接](#)窗口。

步骤5 根据实际情况选择大数据云服务，单击“[下一步](#)”，进入[详细配置](#)页面。

步骤6 根据对应的参数配置说明，配置云服务参数。

- [DLI服务参数配置说明](#)

- [CloudTable\(ClickHouse\)服务参数配置说明](#)
- [CloudTable\(HBase\)服务参数配置说明](#)

表 2-83 DLI 服务参数配置说明

参数	配置说明
所属源	选择目的端。
连接名称	默认创建名称为“DLI-4位随机数（字母和数字）”，也可以自定义名称。
DLI凭证	选择DLI云服务的凭证。凭证的添加方法请参考 添加资源凭证 中的大数据-DLI。如果所选凭证与登录迁移中心账号凭证相同，请勾选“ 所选凭证与登录迁移中心凭证相同 ”，系统将根据所选的区域，自动查询账号在该区域下的项目。
区域/项目	选择需要校验的数据所在区域/项目。
队列名称	输入需要进行校验的DLI队列名称。队列类型必须是SQL队列。

表 2-84 CloudTable(ClickHouse)服务参数配置说明

参数	配置说明
所属源	选择目的端。
连接名称	默认创建名称为“CloudTable-ClickHouse-4位随机数（字母和数字）”，也可以自定义名称。
CloudTable(ClickHouse)凭证	选择CloudTable(ClickHouse)集群的凭证。凭证的添加方法请参考 添加资源凭证 中的大数据-ClickHouse。认证方式选择“ 用户名/密码 ”，填写在CloudTable服务创建的目的端ClickHouse集群的登录用户名和密码。
数据库链接	填写连接到CloudTable(ClickHouse)的数据库链接地址。可以在ClickHouse集群的 基本信息 获取，对应参数为“ 集群访问地址 ”。

表 2-85 CloudTable(HBase)服务参数配置说明

参数	配置说明
所属源	选择目的端。
连接名称	默认创建名称为“CloudTable-HBase-4位随机数（字母和数字）”，也可以自定义名称。
CloudTable(HBase)凭证	非安全集群，无需凭证。

参数	配置说明
安全集群	当前仅支持非安全集群（普通集群）。
数据库链接	填写连接到CloudTable(HBase)的数据库链接地址。可以在HBase集群的 基本信息 获取，对应参数为“ZK链接地址（内网）”。

步骤7 参数配置完成后，单击“[测试连接](#)”，系统会校验配置信息并与云服务建立连接，测试通过代表与云服务连接成功。

步骤8 测试连接通过后，单击“[确认](#)”按钮，创建云服务连接完成。

步骤9 在[连接管理](#)页面，可以查看创建的连接列表以及基本信息。单击操作列的“[修改连接](#)”，可以修改连接参数配置。

----结束

2.12.5.3 创建目的端组件连接

对大数据组件数据进行一致性校验时，需要与大数据组件建立连接，用于数据查询。

目前支持创建连接的大数据组件包括：

- Doris
- HBase
- ClickHouse
- Hive Metastore

操作步骤

步骤1 登录[迁移中心](#)管理控制台。

步骤2 在左侧导航栏选择“[实施>大数据校验](#)”，进入[大数据校验](#)页面。在页面左上角的当前项目下拉列表中选择[迁移项目](#)。

步骤3 在功能入口区域，单击“[连接管理](#)”，进入[连接管理](#)页面。

步骤4 单击页面右上角的“[新增连接](#)”，右侧弹出[新增连接](#)窗口。

步骤5 根据实际情况选择大数据组件，单击“[下一步](#)”，进入[详细配置](#)页面。

步骤6 根据对应的参数配置说明，配置组件连接参数。

- [Doris组件参数配置说明](#)
- [HBase组件参数配置说明](#)
- [ClickHouse组件参数配置说明](#)
- [Hive Metastore连接参数配置说明](#)

表 2-86 Doris 组件参数配置说明

参数	配置说明
所属源	选择目的端。
连接名称	默认创建名称为“Doris-4位随机数（字母和数字）”，也可以自定义名称。
Doris凭证	选择在Edge添加的目的端Doris凭证。凭证的添加方式请参考 添加资源凭证 ，类型选择“大数据-Doris”。
数据库地址	填写目的端Doris集群部署的服务器IP地址。 可登录 FusionInsight Manager ，选择“集群 > 服务 > Doris”，查看“ Leader所在的主机 ”获取。
数据库端口	填写目的端Doris集群部署的服务器的端口。 可登录 FusionInsight Manager ，选择“集群 > 服务 > Doris > 配置”，搜索“ query_port ”获取。
数据库名	填写目的端Doris数据库名称。

表 2-87 HBase 组件参数配置说明

参数	配置说明
所属源	选择目的端。
连接名称	默认创建名称为“HBase-4位随机数（字母和数字）”，也可以自定义名称。
HBase凭证	选择在Edge添加的目的端HBase凭证。凭证的添加方式请参考 添加资源凭证 ，类型选择“大数据-HBase”。
安全集群	选择是否部署在安全集群下。
Zookeeper节点IP	填写连接到目的端Zookeeper的IP地址。通常是部署Zookeeper的服务器公网IP地址或内网IP地址。
Zookeeppeer连接端口	填写连接到目的端Zookeeppeer的端口。默认为2181。
HBase 版本	选择目的端HBase版本。

表 2-88 ClickHouse 组件参数配置说明

参数	配置说明
所属源	选择目的端。
连接名称	默认创建名称为“ClickHouse-4位随机数（字母和数字）”，也可以自定义名称。

参数	配置说明
ClickHouse凭证（可选）	选择在Edge添加的目的端ClickHouse凭证。凭证的添加方法请参考 添加资源凭证 中的大数据-ClickHouse。
安全集群	选择是否部署在安全集群下。
ClickHouse Server IP	填写连接到目的端ClickHouse服务器的IP地址。通常是运行ClickHouse的主机IP地址。
HTTP端口	非安全集群下，填写与目的端ClickHouse 服务器通信的HTTP 端口。默认为8123。
HTTP SSL/TLS端口	安全集群下，填写与目的端ClickHouse 服务器通信的HTTPS 端口。

表 2-89 Hive Metastore 连接参数配置说明

参数	配置说明
所属源	选择目的端。
连接名称	默认创建名称为“Hive-Metastore-4位随机数（字母和数字）”，也可以自定义名称。
启用安全连接	选择是否启用安全连接。 <ul style="list-style-type: none"> MetaStore在非安全集群下，可以不启用安全连接。 MetaStore在安全集群下，需要启用安全连接并选择凭证。凭证的获取以及添加方法请参考添加资源凭证中的大数据-Hive Metastore。
Hive版本	选择目的端Hive的实际版本。 注意 目前只支持3.x版本。
MetaStore节点IP	填写连接MetaStore的IP地址。
MetaStore Thrift服务端口	填写连接MetaStore的端口，默认为9083。

参数	配置说明
启用元数据仓库连接	<p>在增量数据校验中，如果通过Metastore查询，当分区数量超过30,000时，可能会因为将所有分区信息加载到内存中而导致内存溢出（OOM）。因此，使用MySQL作为元数据仓库可以有效地解决这一问题。</p> <ul style="list-style-type: none"> 不启用：系统通过MetaStore查询Hive表的分区信息。 启用：配置MySQL作为元数据仓库，系统将通过MySQL查询Hive表的分区信息。启用后，需要配置如下参数： <ul style="list-style-type: none"> 元仓类型：目前仅支持MySQL作为元数据仓库。 MySQL凭证：选择访问MySQL数据库的认证信息。认证信息需要添加到Edge然后同步到MgC，方法请参考添加资源凭证。 MySQL节点IP：输入MySQL数据库服务器的IP地址。 MySQL服务端口：输入MySQL数据库服务的端口号。 数据库名：输入存储Hive表元数据的数据库名称。 <p>说明 请确保填写的MySQL凭证、节点IP、服务端口和数据库名与Hive实际使用的MySQL数据库相匹配，否则会导致数据校验失败。</p>

步骤7 参数配置完成后，单击“**测试连接**”，系统会校验配置信息并与组件建立连接，测试通过代表与组件连接成功。

步骤8 测试连接通过后，单击“**确认**”按钮，创建大数据组件连接完成。

步骤9 在**连接管理**页面，可以查看创建的连接列表以及基本信息。单击操作列的“**修改连接**”，可以修改连接参数配置。

----结束

2.12.6 创建统计任务并执行校验

利用创建的源端连接和目的端连接，分别创建源端和目的端统计任务，获取源端和目的端的数据信息，以便于进行数据校验。

支持的大数据组件和统计方式说明请参见[大数据校验概述](#)。

注意事项

- 创建的源端统计任务和目的端统计任务的统计方式必须保持一致。
- HBase安全集群任务与非安全集群任务不能同时执行。HBase安全集群需要加载认证信息，非安全集群连接时需要清除安全集群的认证信息，因此不能同时执行。
- 校验任务不支持跨天执行，执行过程中如果跨0点将导致校验结果异常，请合理规划任务以避免跨天运行。
- 当源端Lindorm服务因欠费被锁定时，虽然可以正常创建数据连接和校验任务，但由于服务锁定限制了数据访问和操作权限，最终会导致校验任务无法正常执行。在进行数据校验前，请确保Lindorm服务未因欠费被锁定，以避免权限问

题。如果服务被锁定，请及时补缴费用以解除锁定。确认服务恢复正常后，重新尝试执行数据校验任务。

约束与限制

- 创建EMR Delta Lake到MRS Delta Lake数据校验任务存在以下约束与限制：
 - 源端EMR环境的Spark版本为3.3.1时，支持Delta Lake（有元数据）和Delta Lake（无元数据）的数据校验。
 - 源端EMR环境的Spark版本为2.4.8时，仅支持Delta Lake（有元数据）的数据校验。
- Hbase校验不支持纯冷数据表。
- 在进行Hive 2.x与Hive 3.x版本间数据校验时，由于Hive 2.x版本在查询固定长度字符串类型char(N)数据时，如果实际数据长度不满足N长度，Hive会以空格补齐至N长度。而Hive 3.x版本在查询时不会进行这种补齐操作。这可能导致在不同版本间比较结果时出现差异。为了避免因版本差异导致的数据不一致问题，建议使用Beeline作为客户端工具进行作业对数。
- 在进行ClickHouse数据校验时，如果源端阿里云的ClickHouse版本为21.8.15.7，目的端华为云的ClickHouse版本为23.3.2.37，由于这两个版本在处理IPv4和IPv6字段类型以及相关函数计算结果上存在差异，因此无法进行字段校验。
- 在执行Hive数据的日级别增量、小时级增量以及指定日期的校验时，如果分区字段为date类型，且分区值的格式不是标准的YYYY-MM-DD时，系统目前不支持对这种场景进行数据校验。

前提条件

- 已[创建表组并添加数据表](#)。
- 已根据组件类型，[创建对应的连接](#)。

操作步骤

步骤1 登录[迁移中心](#)管理控制台。

步骤2 在左侧导航栏选择“[实施>大数据校验](#)”，进入[大数据校验](#)页面。在页面左上角的[当前项目](#)下拉列表中选择[迁移项目](#)。

步骤3 在[功能入口](#)区域，单击“[任务管理](#)”，进入[任务管理](#)页面。

步骤4 单击页面右上角的“[新增任务](#)”，右侧弹出[新增任务](#)窗口。

步骤5 根据实际需求，选择大数据组件类型和统计方式，单击“[下一步](#)”，进入[详细配置](#)页面。

步骤6 根据对应的参数配置说明，配置所选的统计方式参数。

说明

各组件之间配置参数略有区别，请以实际页面为准。

- [全量对数作业参数配置说明](#)
- [日级增量校验参数配置说明](#)
- [小时级增量校验参数配置说明](#)
- [指定日期校验参数配置说明](#)

• 指定时间校验参数配置说明

表 2-90 全量校验参数配置说明

区域	参数	配置说明
基础信息	任务名	默认创建名称为“组件类型全量校验-4位随机数（字母和数字）”，也可以自定义名称。
任务配置	表组	选择需要校验的表所在表组。
	执行机连接	Hive、Delta Lake和Hudi组件配置参数。 <ul style="list-style-type: none"> 创建源端统计任务时，选择源端执行机连接。 创建目的端统计任务时，选择目的端执行机连接。
	DLI连接	数据湖探索（DLI）组件配置参数，创建目的端统计任务。选择 创建的目的端DLI连接 。
	Doris连接	Doris组件配置参数。 <ul style="list-style-type: none"> 创建源端统计任务时，选择源端Doris连接。 创建目的端统计任务时，选择目的端Doris连接。
	HBase连接	HBase组件和CloudTable(HBase)组件配置参数。 <ul style="list-style-type: none"> 创建源端统计任务时，选择源端HBase连接。 创建目的端统计任务时，选择目的端HBase或CloudTable(HBase)连接。
	ClickHouse连接	ClickHouse、阿里云云数据库ClickHouse和CloudTable(Clickhouse)组件配置参数。 <ul style="list-style-type: none"> 创建源端统计任务时，选择源端MRS ClickHouse连接或阿里云云数据库ClickHouse连接。 创建目的端统计任务时，选择目的端MRS ClickHouse或CloudTable(ClickHouse)连接。
	元数据连接（可选）	Delta Lake组件配置参数。 创建Delta Lake统计任务时，可以选择与表组所属集群相关的源端元数据连接，用于提升任务处理效率。
	执行时间	设置作业自动执行时间。当作业激活后，每天都会在设置的时间自动执行。 说明 建议执行时间选择业务的低峰期。

区域	参数	配置说明
	高级选项	<ul style="list-style-type: none"> 并发连接数：进行数据统计时，单个执行机的任务并发数。默认为3，取值范围：1~10。 注意 在创建阿里云Hive校验任务时，请根据源端数据量和EMR Master节点的规格，合理设置并发连接数。建议按照以下两个条件进行设置： <ul style="list-style-type: none"> 条件一：同时运行的任务并发数之和不超过元数据节点总核数的70%。 条件二：任务分配的总资源不超过执行队列资源。计算如下： 内存：num-executors x executor-memory x 并发数 核数：num-executors x executor-cores x 并发数 例如，源端总数据量为500 G，且包含10000张表，其中8张为50 G、10万分区的大表。EMR Master节点规格为8U32G。 <ul style="list-style-type: none"> 根据条件一：最大并发数 = $0.7 \times 8 = 5.6$，向下取整得到最大并发数为5。 根据条件二：需要在后续执行命令参数配置中，选择Spark-sql，并设置如下参数： <div data-bbox="900 965 1171 1149" style="border: 1px solid #ccc; padding: 5px; margin: 10px 0;"> <p>执行命令</p> <p>baseline hive spark-sql</p> <p>executor-memory 4G</p> <p>master yarn</p> <p>num-executors 20</p> <p>executor-cores 2</p> <p>driver-memory 10G 添加参数</p> <p>脚本预览</p> <pre>spark-sql --executor-memory 4G --master yarn --num-executors 20 --executor-cores 2 --driver-memory 10G -f</pre> </div> executor-memory = 4G master = yarn num-executors = 20 executor-cores = 2 driver-memory = 10G 单文件SQL条数：每个需要校验的表会产生一条SQL用以获取数据摘要，每批提交给集群处理的数据文件中所包含的SQL条数。默认为10，推荐取值范围：1~50。 超时时间：每个SQL语句在大数据集群上执行时的最大允许时长，单位为秒（s）。默认为600，推荐取值范围：600~ 7200。 发送SMN通知：选择是否使用SMN消息通知功能，该功能可以通过电子邮件、短信或者自定义URL等方式及时通知您任务状态。 须知 <ul style="list-style-type: none"> 使用该功能前，需要前往SMN控制台创建主题。方法请参考创建SMN主题。 使用该功能可能产生少量费用，具体费用由SMN服务结算。详情请查看SMN计费说明。

区域	参数	配置说明
数据筛选	非分区表处理策略	<p>选择对于非分区表的校验处理策略。</p> <ul style="list-style-type: none"> ● 全部校验：全量处理所有非分区表，不做筛选。 ● 全部忽略：忽略所有非分区表，不做校验。 ● 按更新时间过滤：只处理更新时间在校验范围内的非分区表。注意：非分区表非insert写入业务，元数据更新时间存在偏差，按更新时间过滤分区时可能遗漏分区，慎重选择该选项！
	高级选项	<p>单SQL字段上限：每条SQL查询的表的字段上限。单SQL查询的字段数越多，执行时间越长，如果设置的字段数上限过少，则可能形成更多SQL，影响获取校验结果的效率。默认是0，即无上限，取值范围100~500。</p>
OBS桶配置检查	-	<ul style="list-style-type: none"> ● 如果需要将任务日志文件和内容校验结果上传到OBS桶进行管理和分析，请参考配置OBS桶配置日志桶。配置了日志桶后，会默认使用日志上传功能，将日志数据和内容校验结果上传到指定的OBS桶中。 ● 如果不需要将任务日志文件和内容校验结果上传到OBS桶，请勾选“我确认仅需在Edge侧查看日志和内容校验结果，无需上传至OBS桶”。
命令参数 (Hive配置参数)	安全配置	<p>如果大数据集群启用了安全配置（例如Kerberos安全认证），请勾选此选项，并配置安全认证命令。配置前，需要您手动将包含认证密钥的.keytab文件上传至执行机。</p> <ul style="list-style-type: none"> ● keytab文件地址：输入.keytab文件在执行机的存放地址。 ● keytab文件用户名：输入.keytab文件对应的用户名。
	执行命令	<p>可以通过设置Beeline或Spark-sql命令行参数，对数据进行处理。</p> <ul style="list-style-type: none"> ● Beeline：用于与Hive交互的命令行工具。 ● Spark-sql：用于执行SQL查询和分析数据。
MaxCompute参数 (MaxCompute配置参数)	-	<p>根据实际需求，添加MaxCompute参数。参数详情请参考MaxCompute文档。</p>
统计配置 (DLI配置参数)	参数	<p>根据实际需求进行配置，支持的自定义参数详情请参考自定义参数说明。</p>

区域	参数	配置说明
统计配置 (HBase配置参数)	运行模式	支持以下两种运行模式： Yarn：适用于大型分布式环境，可以充分利用集群资源，提高任务的并行度和效率。 Local：适用于小规模数据集或开发测试环境，可以快速进行调试和验证。
	参数	根据所选的运行模式和实际需求，添加命令行参数。
命令参数 (Delta Lake和Hudi配置参数)	安全配置 (仅创建Delta Lake目的端任务时可以配置)	如果大数据集群启用了安全配置 (例如Kerberos安全认证)，请勾选此选项，并配置安全认证命令。配置前，需要您手动将包含认证密钥的 .keytab 文件上传至执行机。 <ul style="list-style-type: none"> • keytab文件地址：输入.keytab文件在执行机的存放地址。 • keytab文件用户名：输入.keytab文件对应的用户名。
	spark-sql命令	Spark SQL是Spark中用于结构化数据处理的模块。在Spark应用中，可以无缝地使用SQL语句亦或是DataSet API对结构化数据进行查询。更多介绍请参见 SparkSQL原理 。建议保持默认。

区域	参数	配置说明
	spark-submit 命令	<p>通过Spark基本shell命令，提交Spark应用。命令为：</p> <pre>./bin/spark-submit \ --class <main-class> \ --master <master-url> \ ... # other options <application-jar> \ [application-arguments]</pre> <p>参数解释：</p> <ul style="list-style-type: none"> ● --class: Spark应用的类名。 ● --master: Spark用于所连接的master，如 yarn-client, yarn-cluster等。 ● application-jar: Spark应用的jar包的路径。 ● application-arguments: 提交Spark应用的所需要的参数（可以为空）。 <p>以下为特定场景下需要添加的参数配置：</p> <ul style="list-style-type: none"> ● 如果创建的是阿里云 EMR且涉及Spark 3的 Delta Lake源端统计任务，请添加如下参数： <ul style="list-style-type: none"> - 参数名: jars - 参数值: '/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-core_2.12-*.jar,/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-storage-*.jar' <p>注意 参数值需要根据实际环境目录和Delta Lake 的版本进行替换。</p> ● 如果源端是阿里云 EMR Delta 2.1.0版本并且配置了Spark 2.4.8来处理Delta Lake数据，请添加如下参数： <ul style="list-style-type: none"> - 参数名: mgc.delta.spark.version - 参数值: 2

表 2-91 日级增量校验参数配置说明

区域	参数	配置说明
基础信息	任务名	默认创建名称为“组件类型日级增量校验-4位随机数（字母和数字）”，也可以自定义名称。
任务配置	表组	选择需要校验的表所在表组。
	执行机连接	<p>Hive、Delta Lake和Hudi组件配置参数。</p> <ul style="list-style-type: none"> ● 创建源端统计任务时，选择源端执行机连接。 ● 创建目的端统计任务时，选择目的端执行机连接。

区域	参数	配置说明
	DLI连接	数据湖探索 (DLI) 组件配置参数，创建目的端统计任务。选择 创建的DLI连接 。
	Doris连接	Doris组件配置参数。 <ul style="list-style-type: none"> 创建源端统计任务时，选择源端Doris连接。 创建目的端统计任务时，选择目的端Doris连接。
	HBase连接	HBas组件配置参数。 <ul style="list-style-type: none"> 创建源端统计任务时，选择源端HBase连接。 创建目的端统计任务时，选择目的端HBase或CloudTable(HBase)连接。
	元数据连接 (可选)	Hive、Delta Lake组件配置参数。 创建Hive目的端统计任务时可配置，选择Hive目的端元数据连接。用于判断目的端分区是否存在。 创建Delta Lake统计任务时，可以选择与表组所属集群相关的源端元数据连接，用于提升任务处理效率。
	执行时间	设置作业自动执行时间。当作业激活后，每天都会在设置的时间自动执行。

区域	参数	配置说明
	高级选项	<ul style="list-style-type: none"> 并发连接数：进行数据统计时，单个执行机的任务并发数。默认为3，取值范围：1~10。 注意 在创建阿里云Hive校验任务时，请根据源端数据量和EMR Master节点的规格，合理设置并发连接数。建议按照以下两个条件进行设置： <ul style="list-style-type: none"> 条件一：同时运行的任务并发数之和不超过元数据节点总核数的70%。 条件二：任务分配的总资源不超过执行队列资源。计算如下： 内存：num-executors x executor-memory x 并发数 核数：num-executors x executor-cores x 并发数 例如，源端总数据量为500 G，且包含10000张表，其中8张为50 G、10万分区的大表。EMR Master节点规格为8U32G。 <ul style="list-style-type: none"> 根据条件一：最大并发数 = $0.7 \times 8 = 5.6$，向下取整得到最大并发数为5。 根据条件二：需要在后续执行命令参数配置中，选择Spark-sql，并设置如下参数： <div data-bbox="893 958 1165 1120" data-label="Image"> </div> <pre> executor-memory = 4G master = yarn num-executors = 20 executor-cores = 2 driver-memory = 10G </pre> 单文件SQL条数：每个需要校验的表会产生一条SQL用以获取数据摘要，每批提交给集群处理的数据文件中所包含的SQL条数。默认为10，推荐取值范围：1~50。 超时时间：每个SQL语句在大数据集群上执行时的最大允许时长，单位为秒（s）。默认为600，推荐取值范围：600~ 7200。 发送SMN通知：选择是否使用SMN消息通知功能，该功能可以通过电子邮件、短信或者自定义URL等方式及时通知您任务状态。 须知 <ul style="list-style-type: none"> 使用该功能前，需要前往SMN控制台创建主题。方法请参考创建SMN主题。 使用该功能可能产生少量费用，具体费用由SMN服务结算。详情请查看SMN计费说明。

区域	参数	配置说明
数据筛选	增量区间	<p>选择需要对哪些时间段内的增量数据进行校验。默认统计24小时增量数据。T代表执行时间，T-n代表执行时间前的n x 24小时。</p> <p>勾选统计连续日期，可以选择多个统计时间。</p>
	非分区表处理策略	<p>选择对于非分区表的对数处理策略。</p> <ul style="list-style-type: none"> ● 全部校验：全量处理所有非分区表，不做筛选。 ● 全部忽略：忽略所有非分区表，不做校验。 ● 按更新时间过滤：只处理更新时间在校验范围内的非分区表。注意：非分区表非insert写入业务，元数据更新时间存在偏差，按更新时间过滤分区时可能遗漏分区，慎重选择该选项！
	高级选项	<ul style="list-style-type: none"> ● 分区筛选方式：通过“更新时间”和“创建时间”，判断分区是否属于校验范围内。 <ul style="list-style-type: none"> - 更新时间：指分区内数据最后一次被修改或更新的时间戳。如果您关心的是分区内数据的最新状态或最近一段时间内的数据变更，可以选择“更新时间”作为筛选条件。 - 创建时间：指分区被首次创建时的时间戳。如果您关心的是分区从创建开始到某个时间点内的数据，可以选择“创建时间”作为筛选条件。 ● 分区上限：用于限制查询或处理的分区数量，避免不必要的全表扫描或处理大量不需要的数据。默认为3，取值范围：1~50。 例如，设置分区上限为3，系统只处理降序排序后分区列表中的前3个分区。 ● 单SQL字段上限：每条SQL查询的表的字段上限。单SQL查询的字段数越多，执行时间越长，如果设置的字段数上限过少，则可能形成更多SQL，影响获取校验结果的效率。默认是0，即无上限，取值范围100~500。
OBS桶配置检查	-	<ul style="list-style-type: none"> ● 如果需要将任务日志文件和内容校验结果上传到OBS桶进行管理和分析，请参考配置OBS桶配置日志桶。配置了日志桶后，会默认使用日志上传功能，将日志数据和内容校验结果上传到指定的OBS桶中。 ● 如果不需要将任务日志文件和内容校验结果上传到OBS桶，请勾选“我确认仅需在Edge侧查看日志和内容校验结果，无需上传至OBS桶”。

区域	参数	配置说明
执行脚本 (Hive配置参数)	安全配置	<p>如果大数据集群启用了安全配置（例如Kerberos安全认证），请勾选此选项，并配置安全认证命令。配置前，需要您手动将包含认证密钥的.keytab文件上传至执行机。</p> <ul style="list-style-type: none"> keytab文件地址：输入.keytab文件在执行机的存放地址。 keytab文件用户名：输入.keytab文件对应的用户名。
	执行命令	<p>可以通过设置Beeline或Spark-sql命令行参数，对数据进行处理。</p> <ul style="list-style-type: none"> Beeline：用于与Hive交互的命令行工具。 Spark-sql：用于执行SQL查询和分析数据。
MaxCompute参数 (MaxCompute配置参数)	-	根据实际需求，添加MaxCompute参数。参数详情请参考 MaxCompute文档 。
统计配置 (DLI配置参数)	参数	根据实际需求进行配置，支持的自定义参数详情请参考 自定义参数说明 。
统计配置 (HBase配置参数)	运行模式	<p>支持以下两种运行模式：</p> <p>Yarn：适用于大型分布式环境，可以充分利用集群资源，提高任务的并行度和效率。</p> <p>Local：适用于小规模数据集或开发测试环境，可以快速进行调试和验证。</p>
	参数	根据所选的运行模式和实际需求，添加命令行参数。
命令参数 (Delta Lake和Hudi配置参数)	安全配置（仅创建Delta Lake目的端任务时可以配置）	<p>如果大数据集群启用了安全配置（例如Kerberos安全认证），请勾选此选项，并配置安全认证命令。配置前，需要您手动将包含认证密钥的.keytab文件上传至执行机。</p> <ul style="list-style-type: none"> keytab文件地址：输入.keytab文件在执行机的存放地址。 keytab文件用户名：输入.keytab文件对应的用户名。
	spark-sql命令	Spark SQL是Spark中用于结构化数据处理的模块。在Spark应用中，可以无缝地使用SQL语句亦或是DataSet API对结构化数据进行查询。更多介绍请参见 SparkSQL原理 。建议保持默认。

区域	参数	配置说明
	spark-submit 命令	<p>通过Spark基本shell命令，提交Spark应用。命令为：</p> <pre>./bin/spark-submit \ --class <main-class> \ --master <master-url> \ ... # other options <application-jar> \ [application-arguments]</pre> <p>参数解释：</p> <ul style="list-style-type: none"> ● --class: Spark应用的类名。 ● --master: Spark用于所连接的master，如yarn-client, yarn-cluster等。 ● application-jar: Spark应用的jar包的路径。 ● application-arguments: 提交Spark应用的所需要的参数（可以为空）。 <p>以下为特定场景下需要添加的参数配置：</p> <ul style="list-style-type: none"> ● 如果创建的是阿里云 EMR且涉及Spark 3的Delta Lake源端统计任务，请添加如下参数： <ul style="list-style-type: none"> - 参数名: jars - 参数值: '/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-core_2.12-*.jar,/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-storage-*.jar' <p>注意 参数值需要根据实际环境目录和Delta Lake 的版本进行替换。</p> ● 如果源端是阿里云 EMR Delta 2.1.0版本并且配置了Spark 2.4.8来处理Delta Lake数据，请添加如下参数： <ul style="list-style-type: none"> - 参数名: mgc.delta.spark.version - 参数值: 2

表 2-92 小时级增量校验参数配置说明

区域	参数	配置说明
基础信息	任务名	默认创建名称为“组件类型小时级增量校验-4位随机数（字母和数字）”，也可以自定义名称。
任务配置	表组	选择需要校验的表所在表组。
	执行机连接	<p>Hive、Delta Lake和Hudi组件配置参数。</p> <ul style="list-style-type: none"> ● 创建源端统计任务时，选择源端执行机连接。 ● 创建目的端统计任务时，选择目的端执行机连接。

区域	参数	配置说明
	DLI连接	数据湖探索（DLI）组件配置参数，创建目的端统计任务。选择 创建的DLI连接 。
	Doris连接	Doris组件配置参数。 <ul style="list-style-type: none"> 创建源端统计任务时，选择源端Doris连接。 创建目的端统计任务时，选择目的端Doris连接。
	元数据连接 (可选)	Hive、Delta Lake组件配置参数。 创建Hive目的端统计任务时可配置，选择Hive目的端元数据连接。用于判断目的端分区是否存在。 创建Delta Lake统计任务时，可以选择与表组所属集群相关的源端元数据连接，用于提升任务处理效率。
	开始时间	设置作业自动执行时间。当作业激活后，每天都会在设置的时间自动执行。

区域	参数	配置说明
	高级选项	<ul style="list-style-type: none"> 并发连接数：进行数据统计时，单个执行机的任务并发数。默认为3，取值范围：1~10。 注意 在创建阿里云Hive校验任务时，请根据源端数据量和EMR Master节点的规格，合理设置并发连接数。建议按照以下两个条件进行设置： <ul style="list-style-type: none"> 条件一：同时运行的任务并发数之和不超过元数据节点总核数的70%。 条件二：任务分配的总资源不超过执行队列资源。计算如下： 内存：num-executors x executor-memory x 并发数 核数：num-executors x executor-cores x 并发数 例如，源端总数据量为500 G，且包含10000张表，其中8张为50 G、10万分区的大表。EMR Master节点规格为8U32G。 <ul style="list-style-type: none"> 根据条件一：最大并发数 = 0.7×8=5.6，向下取整得到最大并发数为5。 根据条件二：需要在后续执行命令参数配置中，选择Spark-sql，并设置如下参数：  <pre> executor-memory = 4G master = yarn num-executors = 20 executor-cores = 2 driver-memory = 10G </pre> <ul style="list-style-type: none"> 单文件SQL条数：每个需要校验的表会产生一条SQL用以获取数据摘要，每批提交给集群处理的数据文件中所包含的SQL条数。默认为10，推荐取值范围：1~50。 超时时间：每个SQL语句在大数据集群上执行时的最大允许时长，单位为秒（s）。默认为600，推荐取值范围：600~ 7200。 发送SMN通知：选择是否使用SMN消息通知功能，该功能可以通过电子邮件、短信或者自定义URL等方式及时通知您任务状态。 须知 <ul style="list-style-type: none"> 使用该功能前，需要前往SMN控制台创建主题。方法请参考创建SMN主题。 使用该功能可能产生少量费用，具体费用由SMN服务结算。详情请查看SMN计费说明。
数据筛选	执行周期	设置任务的执行频率和持续时间。

区域	参数	配置说明
	非分区表处理策略	<p>选择对于非分区表的校验处理策略。</p> <ul style="list-style-type: none"> ● 全部校验：全量处理所有非分区表，不做筛选。 ● 全部忽略：忽略所有非分区表，不做校验。 ● 按更新时间过滤：只处理更新时间在校验范围内的非分区表。注意：非分区表非insert写入业务，元数据更新时间存在偏差，按更新时间过滤分区时可能遗漏分区，慎重选择该选项！
	高级选项	<ul style="list-style-type: none"> ● 分区筛选方式：通过“更新时间”和“创建时间”，判断分区是否属于校验范围内。 <ul style="list-style-type: none"> - 更新时间：指分区内数据最后一次被修改或更新的时间戳。如果您关心的是分区内数据的最新状态或最近一段时间内的数据变更，可以选择“更新时间”作为筛选条件。 - 创建时间：指分区被首次创建时的时间戳。如果您关心的是分区从创建开始到某个时间点内的数据，可以选择“创建时间”作为筛选条件。 ● 分区上限：用于限制查询或处理的分区数量，避免不必要的全表扫描或处理大量不需要的数据。默认为3，取值范围：1~50。 例如，设置分区上限为3，系统只处理降序排序后分区列表中的前3个分区。 ● 单SQL字段上限：每条SQL查询的表的字段上限。单SQL查询的字段数越多，执行时间越长，如果设置的字段数上限过少，则可能形成更多SQL，影响获取校验结果的效率。默认是0，即无上限，取值范围100~500。
OBS桶配置检查	-	<ul style="list-style-type: none"> ● 如果需要将任务日志文件和内容校验结果上传到OBS桶进行管理和分析，请参考配置OBS桶配置日志桶。配置了日志桶后，会默认使用日志上传功能，将日志数据和内容校验结果上传到指定的OBS桶中。 ● 如果不需要将任务日志文件和内容校验结果上传到OBS桶，请勾选“我确认仅需在Edge侧查看日志和内容校验结果，无需上传至OBS桶”。
执行脚本（Hive配置参数）	安全配置	<p>如果大数据集群启用了安全配置（例如Kerberos安全认证），请勾选此选项，并配置安全认证命令。配置前，需要您手动将包含认证密钥的<code>.keytab</code>文件上传至执行机。</p> <ul style="list-style-type: none"> ● keytab文件地址：输入<code>.keytab</code>文件在执行机的存放地址。 ● keytab文件用户名：输入<code>.keytab</code>文件对应的用户名。

区域	参数	配置说明
	执行命令	<p>可以通过设置Beeline或Spark-sql命令行参数，对数据进行处理。</p> <ul style="list-style-type: none"> • Beeline: 用于与 Hive交互的命令行工具。 • Spark-sql: 用于执行 SQL 查询和分析数据。
MaxCompute参数 (MaxCompute组件配置参数)	-	根据实际需求，添加MaxCompute参数。参数详情请参考 MaxCompute文档 。
统计配置 (DLI配置参数)	参数	根据实际需求进行配置，支持的自定义参数详情请参考 自定义参数说明 。
命令参数 (Delta Lake和Hudi配置参数)	安全配置 (仅创建Delta Lake目的端任务时可以配置)	<p>如果大数据集群启用了安全配置 (例如Kerberos安全认证)，请勾选此选项，并配置安全认证命令。配置前，需要您手动将包含认证密钥的keytab文件上传至执行机。</p> <ul style="list-style-type: none"> • keytab文件地址: 输入keytab文件在执行机的存放地址。 • keytab文件用户名: 输入keytab文件对应的用户名。
	spark-sql命令	Spark SQL是Spark中用于结构化数据处理的模块。在Spark应用中，可以无缝地使用SQL语句亦或是DataSet API对结构化数据进行查询。更多介绍请参见 SparkSQL原理 。建议保持默认。

区域	参数	配置说明
	spark-submit 命令	<p>通过Spark基本shell命令，提交Spark应用。命令为：</p> <pre>./bin/spark-submit \ --class <main-class> \ --master <master-url> \ ... # other options <application-jar> \ [application-arguments]</pre> <p>参数解释：</p> <ul style="list-style-type: none"> • --class: Spark应用的类名。 • --master: Spark用于所连接的master，如yarn-client, yarn-cluster等。 • application-jar: Spark应用的jar包的路径。 • application-arguments: 提交Spark应用的所需要的参数（可以为空）。 <p>以下为特定场景下需要添加的参数配置：</p> <ul style="list-style-type: none"> • 如果创建的是阿里云 EMR且涉及Spark 3的Delta Lake源端统计任务，请添加如下参数： <ul style="list-style-type: none"> - 参数名: jars - 参数值: '/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-core_2.12-*.jar,/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-storage-*.jar' <p>注意 参数值需要根据实际环境目录和Delta Lake 的版本进行替换。</p> • 如果源端是阿里云 EMR Delta 2.1.0版本并且配置了Spark 2.4.8来处理Delta Lake数据，请添加如下参数： <ul style="list-style-type: none"> - 参数名: mgc.delta.spark.version - 参数值: 2

表 2-93 指定日期校验参数配置说明

区域	参数	配置说明
基础信息	任务名	默认创建名称为“组件类型指定日期校验-4位随机数（字母和数字）”，也可以自定义名称。
任务配置	表组	选择需要校验的表所在表组。
	执行机连接	<p>Hive、Delta Lake和Hudi组件配置参数。</p> <ul style="list-style-type: none"> • 创建源端统计任务时，选择源端执行机连接。 • 创建目的端统计任务时，选择目的端执行机连接。

区域	参数	配置说明
	DLI连接	数据湖探索 (DLI) 组件配置参数，创建目的端统计任务。选择 创建的DLI连接 。
	元数据连接 (可选)	Hive、Delta Lake组件配置参数。 创建Hive目的端统计任务时可配置，选择Hive目的端元数据连接。用于判断目的端分区是否存在。 创建Delta Lake统计任务时，可以选择与表组所属集群相关的源端元数据连接，用于提升任务处理效率。
	执行时间	设置作业自动执行时间。当作业激活后，每天都会在设置的时间自动执行。

区域	参数	配置说明
	高级选项	<ul style="list-style-type: none"> 并发连接数：进行数据统计时，单个执行机的任务并发数。默认为3，取值范围：1~10。 注意 在创建阿里云Hive校验任务时，请根据源端数据量和EMR Master节点的规格，合理设置并发连接数。建议按照以下两个条件进行设置： <ul style="list-style-type: none"> 条件一：同时运行的任务并发数之和不超过元数据节点总核数的70%。 条件二：任务分配的总资源不超过执行队列资源。计算如下： 内存：num-executors x executor-memory x 并发数 核数：num-executors x executor-cores x 并发数 例如，源端总数据量为500 G，且包含10000张表，其中8张为50 G、10万分区的大表。EMR Master节点规格为8U32G。 <ul style="list-style-type: none"> 根据条件一：最大并发数 = $0.7 \times 8 = 5.6$，向下取整得到最大并发数为5。 根据条件二：需要在后续执行命令参数配置中，选择Spark-sql，并设置如下参数： <div data-bbox="893 958 1165 1131" data-label="Form"> <p>执行命令</p> <p>baseline hive spark-sql</p> <p>executor-memory 4G</p> <p>master yarn</p> <p>num-executors 20</p> <p>executor-cores 2</p> <p>driver-memory 10G 添加参数</p> <p>脚本预览</p> <pre>spark-sql --executor-memory 4G --master yarn --num-executors 20 --executor-cores 2 --driver-memory 10G -f</pre> <p>executor-memory = 4G master = yarn num-executors = 20 executor-cores = 2 driver-memory = 10G</p> </div> 单文件SQL条数：每个需要校验的表会产生一条SQL用以获取数据摘要，每批提交给集群处理的数据文件中所包含的SQL条数。默认为10，推荐取值范围：1~50。 超时时间：每个SQL语句在大数据集群上执行时的最大允许时长，单位为秒（s）。默认为600，推荐取值范围：600~ 7200。 发送SMN通知：选择是否使用SMN消息通知功能，该功能可以通过电子邮件、短信或者自定义URL等方式及时通知您任务状态。 须知 <ul style="list-style-type: none"> 使用该功能前，需要前往SMN控制台创建主题。方法请参考创建SMN主题。 使用该功能可能产生少量费用，具体费用由SMN服务结算。详情请查看SMN计费说明。

区域	参数	配置说明
数据筛选	指定日期	<p>选择需要对哪些时间段内的增量数据进行校验。默认统计24小时增量数据。T代表执行时间，T-n代表执行时间前的n x 24小时。</p> <p>勾选统计连续日期，可以选择多个统计时间。</p>
	非分区表处理策略	<p>选择对于非分区表的对数处理策略。</p> <ul style="list-style-type: none"> ● 全部校验：全量处理所有非分区表，不做筛选。 ● 全部忽略：忽略所有非分区表，不做校验。 ● 按更新时间过滤：只处理更新时间在校验范围内的非分区表。注意：非分区表非insert写入业务，元数据更新时间存在偏差，按更新时间过滤分区时可能遗漏分区，慎重选择该选项！
	高级选项	<p>单SQL字段上限：每条SQL查询的表的字段上限。单SQL查询的字段数越多，执行时间越长，如果设置的字段数上限过少，则可能形成更多SQL，影响获取校验结果的效率。默认是0，即无上限，取值范围100~500。</p>
OBS桶配置检查	-	<ul style="list-style-type: none"> ● 如果需要将任务日志文件和内容校验结果上传到OBS桶进行管理和分析，请参考配置OBS桶配置日志桶。配置了日志桶后，会默认使用日志上传功能，将日志数据和内容校验结果上传到指定的OBS桶中。 ● 如果不需要将任务日志文件和内容校验结果上传到OBS桶，请勾选“我确认仅需在Edge侧查看日志和内容校验结果，无需上传至OBS桶”。
执行脚本 (Hive配置参数)	安全配置	<p>如果大数据集群启用了安全配置（例如Kerberos安全认证），请勾选此选项，并配置安全认证命令。配置前，需要您手动将包含认证密钥的keytab文件上传至执行机。</p> <ul style="list-style-type: none"> ● keytab文件地址：输入keytab文件在执行机的存放地址。 ● keytab文件用户名：输入keytab文件对应的用户名。
	执行命令	<p>可以通过设置Beeline或Spark-sql命令行参数，对数据进行处理。</p> <ul style="list-style-type: none"> ● Beeline：用于与Hive交互的命令行工具。 ● Spark-sql：用于执行SQL查询和分析数据。
MaxCompute参数 (MaxCompute配置参数)	-	<p>根据实际需求，添加MaxCompute参数。参数详情请参考MaxCompute文档。</p>

区域	参数	配置说明
统计配置 (DLI配置参数)	参数	根据实际需求进行配置，支持的自定义参数详情请参考 自定义参数说明 。
命令参数 (Delta Lake 和Hudi配置 参数)	安全配置 (仅 创建Delta Lake目的端任 务时可以配 置)	<p>如果大数据集群启用了安全配置 (例如Kerberos安全认证)，请勾选此选项，并配置安全认证命令。配置前，需要您手动将包含认证密钥的<code>.keytab</code>文件上传至执行机。</p> <ul style="list-style-type: none"> • keytab文件地址：输入<code>.keytab</code>文件在执行机的存放地址。 • keytab文件用户名：输入<code>.keytab</code>文件对应的用户名。
	spark-sql命令	Spark SQL是Spark中用于结构化数据处理的模块。在Spark应用中，可以无缝地使用SQL语句亦或是DataSet API对结构化数据进行查询。更多介绍请参见 SparkSQL原理 。建议保持默认。

区域	参数	配置说明
	spark-submit 命令	<p>通过Spark基本shell命令，提交Spark应用。命令为：</p> <pre>./bin/spark-submit \ --class <main-class> \ --master <master-url> \ ... # other options <application-jar> \ [application-arguments]</pre> <p>参数解释：</p> <ul style="list-style-type: none"> ● --class: Spark应用的类名。 ● --master: Spark用于所连接的master，如yarn-client, yarn-cluster等。 ● application-jar: Spark应用的jar包的路径。 ● application-arguments: 提交Spark应用的所需要的参数（可以为空）。 <p>以下为特定场景下需要添加的参数配置：</p> <ul style="list-style-type: none"> ● 如果创建的是阿里云 EMR且涉及Spark 3的Delta Lake源端统计任务，请添加如下参数： <ul style="list-style-type: none"> - 参数名: jars - 参数值: '/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-core_2.12-*.jar,/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-storage-*.jar' <p>注意 参数值需要根据实际环境目录和Delta Lake 的版本进行替换。</p> ● 如果源端是阿里云 EMR Delta 2.1.0版本并且配置了Spark 2.4.8来处理Delta Lake数据，请添加如下参数： <ul style="list-style-type: none"> - 参数名: mgc.delta.spark.version - 参数值: 2

表 2-94 指定时间校验参数配置说明

区域	参数	配置说明
基础信息	任务名	默认创建名称为“组件类型指定时间校验-4位随机数（字母和数字）”，也可以自定义名称。
任务配置	表组	选择需要校验的表所在表组。
	HBase连接	HBase组件和CloudTable(HBase)组件配置参数。 <ul style="list-style-type: none"> ● 创建源端统计任务时，选择源端HBase连接。 ● 创建目的端统计任务时，选择目的端HBase或CloudTable(HBase)连接。

区域	参数	配置说明
	高级选项	<ul style="list-style-type: none"> 并发连接数: 进行数据统计时, 单个执行机的任务并发数。默认为3, 取值范围: 1~10。 超时时间: 每个SQL语句在大数据集群上执行时的最大允许时长, 单位为秒(s)。默认为600, 推荐取值范围: 600~ 7200。 发送SMN通知: 选择是否使用SMN消息通知功能, 该功能可以通过电子邮件、短信或者自定义URL等方式及时通知您任务状态。 须知 <ul style="list-style-type: none"> 使用该功能前, 需要前往SMN控制台创建主题。方法请参考创建SMN主题。 使用该功能可能产生少量费用, 具体费用由SMN服务结算。详情请查看SMN计费说明。
数据筛选	统计范围	选择需要对哪段时间内的数据进行校验的。
OBS桶配置检查	-	<ul style="list-style-type: none"> 如果需要将任务日志文件和内容校验结果上传到OBS桶进行管理和分析, 请参考配置OBS桶配置日志桶。配置了日志桶后, 会默认使用日志上传功能, 将日志数据和内容校验结果上传到指定的OBS桶中。 如果不需要将任务日志文件和内容校验结果上传到OBS桶, 请勾选“我确认仅需在Edge侧查看日志和内容校验结果, 无需上传至OBS桶”。
统计配置 (HBase配置参数)	运行模式	支持以下两种运行模式: <ul style="list-style-type: none"> YARN 模式: 适用于大型分布式环境, 可以充分利用集群资源, 提高任务的并行度和效率。 Local 模式: 适用于小规模数据集或开发测试环境, 可以快速进行调试和验证。
	参数	根据所选的运行模式和实际需求, 添加命令行参数。

步骤7 配置完成后, 单击“**保存**”按钮, 完成统计任务的创建。创建成功后, 系统会自动将参数配置同步到迁移工具Edge。您可以在**任务列表**查看创建的统计任务及其配置同步状态。

步骤8 配置同步完成后, 您可以选择以下两种方式执行统计任务:

- **自动定时执行**: 按照设置的执行时间, 自动、定时执行作业。
 - 在任务列表, 单击统计任务**调度状态**列的“**激活**”按钮, 弹出**激活**提示窗。
 - 单击“**确认**”按钮, 激活自动化执行作业。
- **手动执行**: 手动立即触发一次作业的执行。
 - 在任务列表, 单击统计任务**操作**列的“**执行作业**”按钮, 弹出**执行作业**提示窗。
 - 单击“**确认**”按钮, 立即执行任务。

步骤9 执行任务后，单击操作列的“**查看实例**”按钮，进入**实例列表**页面。在**实例列表**页面，可以进行如下操作：

- 查看**实例状态**、**进度统计数据**、**作业开始时间和结束时间**等信息。

注意

如果出现校验任务执行时间过长，页面显示不正确的情况，请把客户端执行机 Driver 日志级别设为 ERROR。

- **上传日志**：如果您需要查看作业的执行日志，可以使用**上传日志**功能，将日志文件上传到您的对象存储（OBS）桶中，以便随时查看和分析。使用**上传日志**功能前，需要您先在迁移工具 Edge 上配置日志桶，配置方法请参见**配置日志桶**。
- **取消/终止实例**：如果您需要终止自动定时执行作业或取消正在运行中的作业，可以单击操作列的“**取消/终止实例**”按钮，将执行中的作业取消/终止。
- **补数**：如果存在结果获取失败的表，可以单击**进度统计**列的“**补数**”按钮，对获取对数摘要失败的表进行重新获取。

----结束

2.12.7 查看并导出校验结果

当源端与目的端的统计任务都执行完成后，在**校验结果**页面可以查看并导出任务的执行结果。校验结果将按照任务开始执行的日期进行分批，如果单日多次执行校验任务，系统只会保留最后一次的校验结果，之前的执行结果会被后续的执行所覆盖。

前提条件

已**创建统计任务并执行校验**。

使用须知

当执行求和操作（如 sum 或 allsum）时，如果结果值的位数超过了 Double 类型的有效位数，会导致精度丢失。并且在处理非常大的数值时，由于浮点数的精度限制，可能会导致实际存在的差异因为超出了 Double 类型能表示的有效位数而无法被感知到，从而使得差异率计算结果为 0。

查看校验结果

- 步骤1** 登录**迁移中心**管理控制台。
- 步骤2** 在左侧导航栏选择“**实施>大数据校验**”，进入**大数据校验**页面。在页面左上角的**当前项目**下拉列表中选择**迁移项目**。
- 步骤3** 在功能入口区域，单击“**校验结果**”，进入**校验结果**页面。
- 步骤4** 通过执行日期，筛选出要查看的执行批次。

大数据校验 / 校验结果

< | 校验结果 ▾ ● 迁移工具: 已启用



步骤5 在校验结果列表中，可以进行如下操作：

- 查看校验详情：单击分区达标详情列的“查看”按钮，展开该数据表的作业详情。单击操作列的“查看详情”，进入校验详情页面，查看该数据表的校验详细信息。
- 查看执行历史及状态：单击操作列的“查看历史”，可以查看该表的所有执行批次和达标状态。

----结束

导出校验结果

步骤1 在校验结果页面，单击“导出”按钮，弹出导出校验结果窗口。

步骤2 选择需要导出的批次、表组名以及类型，单击“确认”按钮，系统开始导出校验结果。

步骤3 导出完成后，单击“导出管理”按钮，右侧弹出导出管理窗口。

步骤4 单击操作列的“下载”，可以将对应表的校验结果下载到本地。

----结束

2.12.8 自定义参数说明

在创建数据湖探索 DLI校验任务时，支持的自定义参数及说明参见[表2-95](#)。

表 2-95 DLI 校验任务支持的自定义参数说明

参数名称	默认值	描述
mgc.mc2dli.table.partition.enable	true	DLI表分区为空或者分区不存在时需要通过查询DLI的元数据去判断分区存不存在。 <ul style="list-style-type: none"> • true: 查询DLI表的分区，空分区校验状态为成功，分区不存在状态为失败。 • false: 不查询DLI表的分区，空分区或者分区不存在校验状态都为成功
spark.sql.files.maxRecordsPerFile	0	要写入单个文件的最大记录数。如果该值为零或为负，则没有限制。
spark.sql.autoBroadcastJoinThreshold	209715200	配置执行连接时显示所有工作节点的表的最大字节大小。通过将此值设置为“-1”，可以禁用显示。 说明 当前仅支持运行命令ANALYZE TABLE COMPUTE statistics noscan的配置单元元存储表，和直接根据数据文件计算统计信息的基于文件的数据源表。
spark.sql.shuffle.partitions	200	为连接或聚合过滤数据时使用的默认分区数。

参数名称	默认值	描述
spark.sql.dynamicPartitionOverwrite.enabled	false	<p>当前配置设置为“false”时，DLI在覆盖写之前，会删除所有符合条件的分区。例如，分区表中有一个“2021-01”的分区，当使用INSERT OVERWRITE语句向表中写入“2021-02”这个分区的数据时，会把“2021-01”的分区数据也覆盖掉。</p> <p>当前配置设置为“true”时，DLI不会提前删除分区，而是在运行时覆盖那些有数据写入的分区。</p>
spark.sql.files.maxPartitionBytes	134217728	读取文件时要打包到单个分区中的最大字节数。
spark.sql.badRecordsPath	-	Bad Records的路径。
spark.sql.legacy.correlated.scalar.query.enabled	false	<ul style="list-style-type: none"> ● 该参数设置为true: <ul style="list-style-type: none"> - 当子查询中数据不重复的情况下，执行关联子查询，不需要对子查询的结果去重。 - 当子查询中数据重复的情况下，执行关联子查询，会提示异常，必须对子查询的结果做去重处理，比如max(),min()。 ● 该参数设置为false: <ul style="list-style-type: none"> - 不管子查询中数据重复与否，执行关联子查询时，都需要对子查询的结果去重，比如max(),min()，否则提示异常。